VOLUME 64
NUMBER 8

WHOLE NO. 314
1950

Psychological Monographs:

General and Applied

Combining the *Applied Psychology Monographs* and the *Archives of Psychology* with the *Psychological Monographs*

HERBERT S. CONRAD, *Editor*

The Validity of a Multiple-Choice Projective Test in Psychopathological Screening

By
MARTIN SINGER
Veterans Administration Hospital
Northport, Long Island, N.Y.

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in the University of Pennsylvania

Accepted for publication, November 7, 1949

Price $1.00

Published by
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1515 MASSACHUSETTS AVE., N.W., WASHINGTON 5, D.C.

COPYRIGHT, 1950, BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION

ACKNOWLEDGEMENTS

THE WRITER wishes to express his deep gratitude to Professor Malcolm G. Preston for his invaluable aid in the formulation and execution of this study. Acknowledgements must also be made to Professor Robert A. Brotemarkle, chairman, and to the other members of the department of psychology, University of Pennsylvania, and to Drs. Richard H. Paynter, Morton L. Wadsworth and Earl E. Swartzlander of the Veterans Administration for their interest and help.

The necessary subjects for this study could not have been obtained without the cooperation of very many individuals from various institutions. Space limitations permit only the following token acknowledgements:

Brooklyn State Hospital: Drs. Clarence Bellinger, Nathan Beckenstein, Louis Mantone, James Rappa, Paul Tarantalla, Sidney Tamarin, and Joseph Villara.

Jewish Hospital of Brooklyn (Rehabilitation Clinic): Dr. Joseph L. Abrahamson and Mrs. Eugenia Gorman.

Mt. Sinai Hospital (Veterans Clinic): Dr. Paul Goolker, Mrs. Roslyn Braun, and Miss Ruth Cuker.

Post Graduate Hospital (Veterans Clinic): Dr. William D. Sherwood and Miss Norrie O'Brien.

Veterans Administration: Dr. Helen M. Campbell, Dr. Irwin J. Cohen, Mr. John W. Green, Mr. Robert Lackey, Dr. Bernard Locke, Dr. Harry M. Harter, Dr. James Q. Holsopple, Mr. Thomas MacNamara, Dr. Charles M. Morris, Mr. Martin Munzer, and Dr. Morris J. Tissenbaum.

Finally, thanks are due to Miss Margaret J. Crookhorn and the personnel in the Mental Hygiene Service of the New York Regional Office, Veterans Administration, for their kind assistance.

MARTIN SINGER

CONTENTS

I. Introduction 1

II. Experimental Design 2

III. Description of the Population Tested 7

IV. Treatment of the Data 11

V. Results 14

VI. Discussion of the Results 27

VII. Conclusions 33

VIII. Appendix 34

IX. Bibliography 39

THE VALIDITY OF A MULTIPLE-CHOICE PROJECTIVE TEST IN PSYCHOPATHOLOGICAL SCREENING[1]

I. INTRODUCTION

THERE is marked need for a test which can be rapidly administered and scored, and which has sufficient validity for use as a screening device in the detection of psychopathology. The present study is an attempt to determine whether the projective test approach[2] may be utilized for this purpose.

In the past decade, more and more clinical psychologists have turned to the use of projective tests in their attempts to identify psychopathology. Most of the projective tests have been standardized under conditions of individual administration and free association responses. Such tests have not been suitable as screening instruments since a great deal of time is required by a highly skilled psychologist to administer and interpret the results for each subject. However, there have been some projective tests in which the subjects selected their responses from several multiple choices listed, permitting rapid and objective scoring by untrained personnel.[3] The most widely used test of this type has been the Harrower Multiple Choice Rorschach (19). Inkblots were also employed as stimuli by Terman (52) and by Harrower and Grinker (4, 17). Pictures were used as stimuli by Clark (5), Franck (11), and Horowicz (23), Murray and Stein (38), Szondi (7, 41) and Temple and Amen (8, 51). Single words served as stimuli for Crown (6), Malamud (31), Maller (34), Mittelman and Brodman (36) and Pressey (39, 40), while story plots were used by Kelly and Bishop (27), Taylor (50) and Wolfenstein (59).

A review of the literature indicates that most of the multiple choice projective tests achieved little discrimination. In those cases where a fair amount of validity was reported, faults in the experimental design, such as improper cross validation, use of extreme groups, or testing of too few subjects, made the validity coefficients suspect. On the other hand, none of the studies which found no discrimination was sufficiently comprehensive to warrant a negative conclusion; namely, that the limitations inherent in any attempt to restrict responses to those listed on a test blank preclude valid discrimination with a projective test. Therefore, the present experiment was undertaken in an effort to determine the potentialities of the multiple choice projective test approach in the detection of psychopathology.

A multiple choice projective test was constructed and administered under what were considered "optimum" conditions; i.e., under those conditions which seem most likely to result in the detection of psychopathology. If, with such a test, a fair degree of validity were obtained, then further experimentation with this approach would seem indicated. On the other hand, if very little or no validity were obtained, then it would appear more profitable to explore other approaches—unless, of course, it was established that optimum conditions did not exist so that factors which might lead to radically different results were not included in the present experimental design.

[1] Sponsored by the Veterans Administration and published with the approval of the Chief Medical Director. The statements and conclusions published by the author are a result of his own study and do not necessarily reflect the opinion or policy of the Veterans Administration.

[2] Comprehensive discussions of the projective methods may be found in Frank (12), Sargent (46) and White (57). Each of a series of reviews by Symonds and Samuel (49), Symonds and Krugman (48), and Hertz, Ellis and Symonds (21) has summarized the literature published since the preceding review. Briefly, the projective test approach is defined as one in which the subject is required to structure an ambiguous stimulus. In doing so, he is presumed to reveal his own personality structure.

[3] There have also been some attempts to quantify tests calling for free association responses, by organizing the responses into various categories which are scored linearly; e.g., the Rorschach Inspection Technique as originated by Munroe (37) and modified by Ross and Ross (45). It is felt that the validity of such tests as screening devices should be investigated if a completely objective scoring system has been proven inadequate.

II. EXPERIMENTAL DESIGN

TWO AREAS where optimum conditions must exist may be distinguished: first, in the selection of the particular psychopathology to be identified, and second, in the test methods used to identify this psychopathology.

A. Type of Psychopathology Selected for Identification

In selecting the psychopathology to be identified, it is recognized that the grosser such psychopathology is, the more optimum conditions will be for valid identification. However, the grossness of the psychopathology must be limited so that useful conclusions may be drawn from the results. For example, the successful identification of deteriorated psychotics by a multiple choice projective test would be of little value. A screening device is not needed for this purpose, and no inferences could be drawn as to the probabilities of finer and more useful psychopathological identification. Therefore, the grossness of the psychopathology selected for identification in the present investigation was limited to such distinctions as would be required of a screening device.

Since projective tests have been used most successfully in the detection of psychopathological diagnoses, the attempt was made in the present investigation to differentiate among three diagnostic categories: normals, anxiety reactions, and schizophrenics. These particular categories were selected because they provided adequate distinctions and because their frequency made available a sufficiently large number of subjects in each category.[1] A more detailed consideration of these diagnostic categories will be presented in section III.

B. The Development of a Multiple-Choice Projective Test

The review of previous research with multiple choice projective tests disclosed no test with sufficiently promising results to warrant further exploration of that particular instrument. Therefore, a new test was constructed, based on the results obtained with individual, group, and multiple choice projective tests, and upon recommendations of various authorities in the field of projective testing. In this test, certain conditions were fixed since time considerations limited the length of the test. However, it was possible to vary other factors within the test so that the effect of such variation could be estimated.

1. Factors Not Varied

a. *Stimuli:* The Rorschach inkblots were selected as the stimuli because the Rorschach appears to be the best projective test for the prediction of psychopathology.

According to Rapaport and Schafer, "The use of the Rorschach Test, as one in a battery of tests, in a clinical-psychiatric setting for pur-

[1] It was recognized that specific situations in which a test would be required to discriminate only among subjects in these three diagnostic classifications would rarely arise. However, the purpose of the experiment was *not to develop a test for practical use, but to determine whether the multiple choice projective test approach is capable of valid psychopathological differentiation.* If *potential validity* is established, then efforts can be made to develop a test for practical purposes; e.g., to screen inductees in the armed forces, to select employees in industry, or to discover students with marked personality problems. In order to limit the psychopathological deviation of subjects in this study to that which would exist in populations such as those mentioned above, the most seriously disturbed anxiety reactions and schizophrenics were not tested. (See section III.)

poses of diagnoses and appraisal of adjustment shows . . . . that it is beyond doubt the most potent single diagnostic instrument clinical psychology possesses" (42, p. 73). Brody, in reviewing the literature on mental testing from 1938 to 1942, declared: "Increased use of mental tests for the study of psychoses and allied disorders is conspicuous in recent psychiatric practice. This increase is most prominent first in the use of the Rorschach ink blots" (3, p. 127). Symonds and Krugman point out the superiority of the Rorschach indirectly with their statement: "Next to the Rorschach, the Thematic Apperception Test has claimed most attention. Indeed, in numerous testing programs the Thematic Apperception Test and the Rorschach are listed as the two preferred methods for the study of personality. Actually, much yet remains to be done in exploration of the use of the Thematic Apperception method before it can receive general clinical use" (48, p. 88). Michael and Buhler (35) found the Rorschach to be superior as a diagnostic instrument to a variety of other tests both projective and non-projective.

While in confirmatory experiments (2, 15, 26, 33, 47, 58) poor results were obtained with the Harrower Multiple Choice Rorschach, which used the Rorschach inkblots as stimuli, such results might well have been caused by inadequacies of the specific test rather than by the nature of the stimuli. Hunt and Stevenson have remarked that ". . . . had Harrower-Erickson followed the usual painstaking steps of item analysis and test validation, relying more on empirical investigation and less on a priori judgment in selecting her alternative responses and constructing her scoring key, the results might have been different" (24, p. 110).

b. *Method of administration:* As the purpose of the present study was to construct and evaluate a test which could be rapidly administered, it was thought desirable to follow a group administration procedure. The Rorschach slides were used since various investigators (18, 20, 22, 29, 30) have found that responses made to these slides are comparable to those made to the individual cards. These slides were projected on a screen in a semi-darkened room. An alternative possibility, the use of booklets in which the Rorschach cards were printed, was eliminated because of the expense of printing such booklets, and because the time of stimulus presentation would have been difficult to control, particularly with the non-normal subjects.

c. *The use of multiple choice questions:* It is only with some form of a multiple choice question that the requirement of rapid objective scoring can be met.

d. *The empirical determination of the scoring weights to be given each response:*[2] Results obtained with the Harrower Multiple Choice Rorschach by Due, Wright and Wright (9), Malamud and Malamud (31, 32), Lawshe and Forster (28), and Blair and Hoehn (2) have clearly indicated that a reponse to a multiple choice Rorschach might not have the same psychopathological significance as the same response to the free association Rorschach. Therefore, the psychopathological significance of multiple choice Rorschach responses must be determined independently.

2. *Factors Varied within the Test*

The factors which were varied are summarized below. They will be discussed more fully in subsequent sections.

a. *Structure of the questions:*

(1). *A "relative" versus an "absolute" decision by the subject.* In some questions the subject was required to select one of four percepts (relative decision), while in other questions the subject was required to rate a single percept on a five point rating scale (absolute decision).

(2). *Differences in the types of alternatives*

[2] At this point, it would seem advisable to define several terms which ordinarily may have different connotations.

A *percept* is a possible description of the whole slide or some part of the slide. It may be specific, such as "A bat"; abstract, such as "Death"; or a denial, such as "Nothing at all."

An *alternative* is one of the multiple choices listed in a question. It may be a percept, as when the subject is required to select one of four percepts listed, or it may be a rating, as when the subject is required to select one of five ratings on a rating scale.

A *response* is the answer a subject makes to a question. A subject's response can be either one of the alternatives listed in a question, or it may be a failure to select any alternative.

included in a question. The types of alternatives were varied to form the following question "designs."

(a). Design I (forced "normal" choice) contained what were considered on a priori grounds to be two percepts characteristic of normals, one percept characteristic of anxiety reactions, and one percept characteristic of schizophrenics.[3]

(b). Design II (unlimited choice) contained one "normal" percept, one "anxiety reaction" percept, one "schizophrenic" percept, and one percept worded "Something else."

(c). Design III (forced "abnormal" choice) contained two "anxiety reaction"' percepts and two "schizophrenic" percepts.

b. *Time of stimulus presentation:* As subjects' immediate perceptions might differ from their more reflective perceptions, the time of stimulus presentation included both a brief exposure of the slides and an exposure which gave the subjects as much time as they required.

c. *Length of test:* The total test battery was made long enough so that information might be obtained as to the effect of increasing the size of the test by comparing the results obtained from sections of the test with those obtained from the total test.

d. *Method of determining scoring weights for each response:* While all methods depended upon the relative percentage of responses made by two groups, various formulae were employed which required varying amounts of difference between the two percentages for responses to be given scoring weights.

[3] The diagnostic significance of the various percepts had to be determined on an a priori basis because no empirically established norms were available. Furthermore, it was necessary to base these a priori judgments on the results obtained with the free association Rorschach, since so little was known about multiple choice Rorschach responses.

The writer would like to express his appreciation to Drs. Bruno Klopfer and Zygmunt A. Piotrowski for their assistance in the selection of the percepts believed to be characteristic of the different diagnostic categories. The method by which the different percepts were chosen as characteristic of normals, anxiety reactions, and schizophrenics for the different inkblots is illustrated below by the way in which the "normal" percepts were selected for the first Rorschach blot.

The writer, after consultation with various experienced Rorschach examiners, drew up a list of about 15 percepts which were considered to be the most characteristic of normals. Drs. Klopfer and Piotrowski examined this list independently and each checked off the eight percepts they considered most characteristic. In some cases, they substituted new percepts as better choices than those listed. The writer used the selection of Drs. Klopfer and Piotrowski as a basis for choosing nine percepts for each card. Then, when the structure of a question called for a "normal" percept, one of these nine percepts was used. Which particular one was selected was determined in a relatively random fashion.

3. *The Final Test Battery*

Having determined the psychopathology to be identified (normal, anxiety reaction, and schizophrenia as discussed in section II, A) and the general composition of the test to be used for this identification, the next step was to construct a specific test. After experimentation with preliminary groups, the following test battery, termed the Singer Multiple Choice Rorschach (SMCR) to differentiate it from the Harrower Multiple Choice Rorschach, was developed. (A copy of the SMCR is shown in Appendix A.) The battery consisted of three tests.

a. *Test A:*

(1). *Number of questions:* 30

(2). *Number of percepts listed in a question:* 4

(3). *Type of percepts in a question:* There were 10 questions each of Designs I, II and III. (The three questions for each slide included one of each design. The order in which the designs were presented was varied systematically.)

(4). *Response by the subject:* The subject was required to select one of the four percepts in each question.

(5). *Time of stimulus presentation:* A slide was shown for 10 seconds for each question. Subjects were then allowed a few seconds more in which to record their response if they had not already done so.

b. *Test B:* This test was exactly the same as Test A with the following exceptions:

(1). *Different percepts were used:* However, the diagnostic significance of each percept paralleled that in Test A.

(2). *Time of stimulus presentations:* The subjects were allowed one and one-half minutes to answer the three questions for each slide. Preliminary experimentation had shown that when given an unlimited amount of time, over 95% of the subjects finished within the above time limit.

c. *Test C:*

(1). *Number of questions:* 100

(2). *Number of percepts listed in a question:* 1

(3). *Types of percepts in a question:* There

were 10 questions for each slide, divided as follows:

(a). Four "normal" percepts.
(b). Three "anxiety reaction" percepts.
(c). Three "schizophrenic" percepts, one of which was "Nothing at all."

(4). *Response by the subject:* Each question (percept) was rated by the subject on a five point absolute rating scale as follows:

(a). Exactly like it;
(b). Very much like it;
(c). Somewhat like it;
(d). Only a little like it;
(e). Not at all like it.

The term "like it" referred to how much either the whole slide or any part of it resembled the percepts listed.

(5). *Time of stimulus presentation:*

(a). Slide I. Three minutes.
(b). Slide II. Two and one-half minutes.
(c). Slides III through X. Two minutes.

The decreasing time limits were established because it was found that the subjects required less time to answer a question as they became familiar with the rating technique. Preliminary experimentation indicated that over 95% of the subjects finished within the above time limits when they were allowed unlimited time.

For Slide I subjects were warned when they had one minute left, while for the other slides they were informed when there were thirty seconds remaining.

The differences among the various tests may be observed more clearly in Table 1. The effect of these differences will be discussed in section IV.

4. Administration of the Test

The regular 3¼" by 4¼" Rorschach slides, purchased from the Psychological Corporation, were projected on a screen by a lantern projector illuminated by a lamp of at least 500 watts.

The screen image was generally at least 4′ by 5′ in size. In a few cases the image was slightly smaller because the size of the testing room made it impossible to focus a larger clear image. However, at such times the subjects were seated slightly nearer to the screen than in the cases where the 4′ by 5′ image was employed. It was possible, by having the light enter from the back of the room, to keep the image on the screen quite clear, while enough light existed for the subjects to read the alternatives on their test blanks. The room was kept in this semi-dark condition throughout the test. Subjects were seated at least 12 feet from the screen and at an angle of less than 30° from the center of the screen.

Before each test began, the subjects were given the answer sheets which listed the various possible responses for that test. They were told that slides would be

TABLE 1

CHARACTERISTICS OF THE DIFFERENT TESTS IN THE SMCR

Test	Time of Stimulus Exposure	Response Required of Subjects	Differences in Types of Percepts*						Total
			Design I		Design II		Design III		
			No. of Quest.	Percepts in a Quest.	No. of Quest.	Percepts in a Quest.	No. of Quest.	Percepts in a Quest.	
A	Brief	Select 1 of 4 percepts	10	N-2 AR-1 S-1	10	N-1 AR-1 S-1 "?"-1	10	AR-2 S-2	30
B	Relatively Unlimited	Select 1 of 4 percepts	10	N-2 AR-1 S-1	10	N-1 AR-1 S-1 "?"-1	10	AR-2 S-2	30
C	Relatively Unlimited	Rate each percept on an absolute rating scale.	40	N-1	30	AR-1	30	S-1	100

* N stands for a percept considered on an a priori basis to be characteristic of normals, AR for one characteristic of anxiety reactions, and S for one characteristic of schizophrenics. The "?" stands for the alternative "Something else."

projected on the screen, and that their task was to select a response while the slide was on the screen. No references were made to the Rorschach or to inkblots. The specific procedure for each test was as follows:

a. *Test A.* The first Rorschach slide was projected on the screen for ten seconds, during which time the subjects checked one of the four alternatives in Question 1. The slide was taken off the screen and after a few seconds the same slide was again projected for 10 seconds while the subjects answered Question 2. After an interval of a few seconds, the slide was shown for another 10 seconds while the subject answered Question 3. The second Rorschach slide was then shown for 10 seconds with the subjects responding to Question 4. This procedure continued, with a new slide being used for each three questions, until all 30 questions were answered.

b. *Test B.* The first Rorschach slide was again projected on the screen while the subjects answered Questions 1A, 2A and 3A. The slide remained on the screen for 90 seconds while all three questions were answered. Then the second Rorschach slide was shown for 90 seconds while the subjects responded to the next three questions, and so on until all 30 questions were answered.

c. *Test C.* The first Rorschach slide was shown on the screen for three minutes while the subjects rated all 10 percepts for that slide. The second slide was projected for two and a half minutes while the subjects rated the second 10 percepts. Then the other slides were shown in order for two minutes each while the subjects rated the 10 percepts listed for each slide.

III. DESCRIPTION OF THE POPULATION TESTED

In the original experimental design, it was planned to establish the scoring weight for each response empirically from the results of 150 subjects, consisting of 50 normals, 50 anxiety reactions, and 50 schizophrenics, and to cross validate those weights on a second set of subjects. (The term "set" will always refer to a group consisting of 50 subjects in each of the three diagnostic categories.) However, since it proved impossible to obtain subjects in the various diagnostic categories who were thoroughly evaluated, and since the scoring weights established from only 50 subjects in each category proved highly unstable, a third set of subjects was tested to provide a larger sample upon which to establish scoring weights. After the testing had been completed, it was found possible to add a fourth set consisting of an "excess" of properly qualified subjects who were not needed for the third set, and subjects who had been excluded from previous sets because they did not quite meet all the requirements.

Over 2,000 subjects were tested, as it frequently was necessary, for administrative reasons, to include extraneous subjects in a testing group. The total number of groups tested was 90, of which 39 included only normals, 27 included only hospitalized schizophrenics, and the remaining 24 contained both anxiety reactions and out-patient schizophrenics. The results of 653 subjects, 250 normals, 203 anxiety reactions, and 200 schizophrenics, were used.

A. Age and Education

Information was obtained as to the age and education of the subjects, and the three diagnostic groups proved to be fairly homogeneous in these respects. The mean age of the normals was 25.0 years, that of the anxiety reactions was 30.0 years, and of the schizophrenics 28.4 years. Standard Deviations were 5.0, 6.6 and 6.4 respectively. There was a great deal of overlap among the groups with 93% of the normals, 87% of the anxiety reactions, and 87% of the schizophrenics falling between the ages of 20 and 35.

The educational level of the subjects was determined by the highest grade completed in school. Information was lacking for 5 normals, 6 anxiety reactions, and 21 schizophrenics, but there was no reason to believe that the educational level of these subjects was much different from that of the others. The mean highest grades reached were 13.2 for the normals, 11.4 for the anxiety reactions and 11.5 for the schizophrenics, with respective Standard Deviations of 2.0, 2.5 and 2.6. Seventy-eight per cent of the normals, 45% of the anxiety reactions, and 46% of the schizophrenics were graduated from high school. The percentages for those completing the second year of high school were 96%, 71% and 72% respectively. The few subjects who had less than an eighth grade education (6% of the anxiety reactions and 7% of the schizophrenics) were generally older or foreign born individuals.

No attempt was made to equate educational level, since the lower scholastic accomplishment is characteristic of the general anxiety reaction and schizophrenic population. It was felt that there was sufficient overlap among the three diagnostic groups to preclude any substantial differences arising merely as a reflection of variations in education. The possible effect of differences in age and amount of schooling has been evaluated in section V,B,8.

B. Description of Subjects in Each Diagnostic Category

The general characteristics of the subjects in each diagnostic category will be discussed first; then the composition of the four different sets will be presented.

1. *Normals*

Subjects in this group were drawn from veterans who came to the New York Regional Office of the Veterans Ad-

ministration for a wide variety of non-psychiatric reasons (e.g., to obtain information relating to insurance, benefits under the G.I. Bill of Rights, medical or dental treatment, etc.). All such veterans were first required to visit the Contact Division where they were interviewed and then routed to the appropriate section of the VA. Since the flow of veterans varied greatly during the week, at peak periods it was necessary for them to wait an hour or more before being interviewed. At this time, an announcement was made over a loud speaker asking male veterans of World War II to volunteer to take a personality test while they were waiting. All such veterans who took the test were identified merely by a number which they placed on each page of their test blank. This was done to insure honesty of response as much as possible.

Subjects were selected as normals on the basis of their responses to two questionnaires, one of the Cornell Selectee Index, Form N (53, 54, 55, 56), and the other a six-item questionnaire (listed in Appendix B) which was validated[1] against the Cornell Index, in an effort to include the best adjusted individuals in this group.

2. *Anxiety Reactions*

Subjects in this category were obtained from the VA Mental Hygiene Services of New York, Brooklyn and Philadelphia, and from the Rehabilitation Clinic of the Jewish Hospital of Brooklyn, the Mt. Sinai Hospital Veterans Clinic, and the Post Graduate Hospital Veterans Clinic. All subjects were out-patients and veterans of World War II.

Originally, only patients diagnosed by a psychiatrist who had treated them for at least several months were accepted in this group. However, it became impossible to obtain a sufficient number of subjects fulfilling this criterion. As the research progressed, there was a great decrease in the number of veterans diagnosed as anxiety reactions since, according to several psychiatrists consulted, those patients whose maladjustment arose primarily out of situational difficulties in the armed service or in becoming readjusted to civilian life either recovered or "bound" their anxiety in a specific neurotic or psychotic reaction pattern. Many patients who retained their free floating anxiety would be diagnosed by a psychiatrist, after a few months of treatment, in terms of the dynamic factors underlying the anxiety rather than as anxiety reactions.

Finally, for those patients still diagnosed as anxiety reactions, interrogation of the psychiatrists making these diagnoses generally revealed that this category was being used as a "waste basket" and that many of the patients had either a mixed psychoneurosis, or were possible latent schizophrenics.

Therefore, in order to secure the desired number of subjects, it became necessary to accept a much less rigid criterion; namely, the diagnosis of a screening psychiatrist who had interviewed the patient for about 45 minutes. It is recognized that many of the patients diagnosed by a screening psychiatrist as anxiety reactions might, after more intensive diagnostic work, be classified differently. However, none of these subjects may be considered normal in the sense of being well adjusted, relatively few could have been schizophrenic, and

[1] Only 10 of 233 subjects who answered all "No's" to this questionnaire, which was the normality criterion, scored over 14 on the Cornell Selectee Index, Form N, and only 34 of the 233 scored over 10.

all had at least a fair amount of free floating anxiety.

3. *Schizophrenics*

About half the subjects in this group came from the Veterans Hospital at Northport, while most of the others, some of whom were also veterans of World War II, came from Brooklyn State Hospital. A small number were being treated as out-patients in the New York and Brooklyn Mental Hygiene Services of the VA.

For the most part, the subjects were in relatively good condition. Not only were the best patients selected from the two hospitals, but the patients had to be willing and able to fill out a fairly complicated test satisfactorily. ("Satisfactory" standards are described in the footnote in Section IV,B,2.) No coercion was used to make the subjects take the test, and even less aid was given them than under many group testing situations, since no proctor assistance was made available during the test.

Although question concerning the original diagnosis of schizophrenia existed in less than a half dozen subjects, many of the patients were considered at the time of testing to be at or near remission. Almost half the Northport subjects were open ward patients, and many were released or sent out on trial visits less than two weeks after taking the test. Thus, subjects in this category represented the "best" schizophrenic population; that is, the type of subject who often might not be identified in a brief interview or from responses to a questionnaire.[2]

[2] A survey of the 110 Northport subjects, made from one to two years after testing, revealed that 33 had been discharged, 17 were out on trial visit status, 2 were AWOL (absent without leave), 5 had been transferred to other hospitals, and 53 remained. Thus, about half no longer required hospitalization. Of those remaining, many have been in good enough condition to be sent out on trial visit or leave of absence since their testing.

No attempt was made to use the schizophrenic sub-classifications (catatonic, hebephrenic, simple, and paranoid) because nearly all psychiatrists felt that the sub-classifications they gave to many of the subjects were unreliable.

4. *Special Characteristics of the Four Sets of Subjects*

a. *Set C1.* While each set of subjects was supposed to include 50 persons in each of the three diagnostic categories, there were actually 100 normals and 53 anxiety reactions and 50 schizophrenics in this set because some of the subjects did not complete the entire battery. However, the combined results of the subjects in each category was equivalent to results which would have been obtained had there been 50 subjects in each category completing the test battery.

(1). *Normals.* These subjects were available for but half an hour. Therefore, they were screened with only the six-item questionnaire, and 50 were given Tests A and B while another 50 were given Test C.

(2). *Anxiety Reactions.* Most of these subjects were thoroughly evaluated by a psychiatrist. There were 53 subjects, 41 of whom took all three tests. One subject took only Test A, two took Tests A and B, two took Tests A and C, and seven took Tests B and C. These discrepancies occurred because some subjects did not answer all the tests satisfactorily, while others came in late and missed part or all of Test A. Results were available for 50 subjects on Tests B and C, but for only 46 subjects on Test A. Therefore, the frequencies in the intervals in the distributions of the last group were multiplied by 50/46 to obtain comparable distributions.

(3). *Schizophrenics.* All 50 subjects took the complete test battery. Forty-two of them were patients at the Northport Veterans Hospital while the other eight were out-patients.

b. *Set C2.* There were 50 subjects in each diagnostic category in this set and in each of the two following sets. All

subjects completed the entire test battery.

(1). *Normals.* These subjects were first screened by the six-item questionnaire, and were given the Cornell Selectee Index, Form N, after completing the SMCR. Only subjects with a score of 10 or below on the Cornell Index were included in this set.

(2). *Anxiety Reactions.* There were 29 subjects selected on the basis of a screening diagnosis. The other 21 subjects were evaluated by the psychiatrist treating them. However, in many cases, the psychiatrist had not seen the patient long enough to formulate more than a tentative diagnostic impression.

(3). *Schizophrenics.* There were 25 subjects from Northport Hospital, and 25 from Brooklyn State Hospital.

c. *Set X.*

(1). *Normals.* These subjects were selected as in Set C2.

(2). *Anxiety Reactions.* There were 41 subjects selected on the basis of a screening diagnosis, while the other 9 subjects were given an evaluation, in some cases a tentative one, by the psychiatrist treating them.

(3). *Schizophrenics.* This group included 25 subjects from Northport, and 25 from Brooklyn State Hospital.

d. *Set C3.* This set was composed of subjects who were in excess after the quota for Set X had been filled, plus subjects who had not quite met the requirements for the previous sets.

(1). *Normals.* There were 10 subjects who met the requirements outlined in C2; 9 subjects who scored 10 or below on the Cornell Index but who had answered one "yes" on the six-item questionnaire; 28 subjects who had scored between 11 and 14 inclusive on the Cornell, 8 of whom answered one "yes" and one of whom answered two "yes's" to the six-item questionnaire; and three subjects who completed only half of the Cornell Index but whose score on this half indicated that they would probably have scored under 10 had they completed the test.

(2). *Anxiety Reactions.* There were 33 "excess" subjects, and 17 subjects who had a supplementary diagnosis, such as "and depression" or "in an inadequate personality" which indicated that they were not relatively pure anxiety reactions. There were 18 in the former group and 11 in the latter whose diagnosis was made by a screening psychiatrist.

(3). *Schizophrenics.* This group included 9 out-patients, 18 open ward patients from Northport Hospital and 23 patients in the process of insulin shock treatment at Brooklyn State Hospital.

Sets C1, C2 and C3 were used as criterion populations upon whom scoring weights were established, while Set X served as the cross validation population.

IV. TREATMENT OF THE DATA

A. Diagnostic Groups Compared

In order to demonstrate that psychopathology can be detected by a multiple choice projective test, it would be necessary to discriminate successfully only between two of the three diagnostic groups tested. Therefore, in the treatment of the data, each pair of two diagnostic categories was considered separately. Thus, a comparison was made between the normals and the anxiety reactions; another comparison was made between the normals and the schizophrenics; and a third comparison was made between the anxiety reactions and the schizophrenics. For convenient reference, each set of two diagnostic groups was termed a "comparison pair," resulting in:

1. A normal-anxiety reaction comparison pair (N-AR)
2. A normal-schizophrenic comparison pair (N-S)
3. An anxiety reaction-schizophrenic comparison pair (AR-S)

Each of these comparison pairs was treated as a distinct unit, and the procedure outlined below for the N-AR pair was duplicated for the other two.

B. Statistical Procedure

1. *Method of Determining Scoring Weights*

The scoring weight given each response was determined by the relative percentage of subjects in each diagnostic group selecting that response. The procedure followed is outlined below.

a. The percentage of normals and anxiety reactions making a response was computed. Possible responses included the various alternatives listed in a question plus the failure to select any alternative.

b. The significance of the difference between these percentages was then determined according to the formulae outlined in paragraph c. below.

(1). If the normals' percentage was significantly higher, the response was given a weight of minus 1.

(2). If the anxiety reactions' percentage was significantly higher, the response was given a weight of plus 1.

(3). If the difference was not significant, the response was weighted 0.

(In the N-S comparison pair, a significantly higher normal percentage received a minus 1 weight, while in the AR-S comparison pair, a significantly higher anxiety reaction percentage received a minus 1 weight.)

c. Two different formulae were used to determine the significance of the difference between the percentages.

(1). The first was a formula adapted from one outlined by Guilford[1] (14, pp. 298-301):

$$W = \frac{P_1 - P_2}{pq},$$

where W = *Weight*. When the weight came to plus 0.5 or above, a weight of plus 1 was given. Similarly, a weight of minus 0.5 or below was weighted minus 1.

P_1 = the percentage of subjects in group 1 making a response.

P_2 = the percentage of subjects in group 2 making a response.

$$p = \frac{P_1 + P_2}{2}$$

$$q = 1 - p$$

(2). The second was the customary formula for finding the significance of a difference between percentages, namely:

$$\text{C.R.} = \frac{P_1 - P_2}{\sqrt{\frac{P_1Q_1}{N_1} + \frac{P_2Q_2}{N_2}}},$$

[1] In Guilford's formula, a constant, plus 4, is added so that possible weights will all be positive and range from 0 through 8. Guilford's formula was modified because the writer found it more convenient to work with plus and minus weights. Weights were restricted to plus or minus 1 as the very few responses which would have received higher weights with the Guilford formula would have caused no appreciable difference in the results.

where C.R. = the critical ratio, and the other symbols have their conventional meaning.

With this formula, the determination of whether a difference is significant will depend upon the size of the critical ratio required. The higher the critical ratio required, the greater the difference between the two percentages will have to be. Four different critical ratios were used as criteria to determine weights as follows:

(a) A C.R. of 1.6449, which is equivalent to a p of .10.
(b) A C.R. of 1.9600, equivalent to a p of .05.
(c) A C.R. of 2.3263, equivalent to a p of .02.
(d) A C.R. of 2.5758, equivalent to a p of .01.

d. Thus, there were five different weighting "systems" used:
(1). That based on the formula of Guilford (referred to in future as weighting system G).
(2). The four based on the regular percentage-difference formula where the p required to indicate a significant difference varied as follows:
(a) $p = .10$ (referred to as weighting system .10).
(b) $p = .05$ (referred to as weighting system .05).
(c) $p = .02$ (referred to as weighting system .02).
(d) $p = .01$ (referred to as weighting system .01).

2. *The Responses Weighted*

Since failure to respond to a stimulus is frequently a psychopathological sign indicative of "stimulus shock" or evasion on the individual Rorschach, such failures were treated as regular responses. Thus, in Tests A and B where the subjects were requested to select one of four percepts, there were five possible responses; four of them were the different percepts listed, while the fifth was the failure to select a percept. Similarly, in Test C, the five steps in the rating scale gave rise to five possible responses, while failure to select any rating provided a sixth possible response. The relatively few cases in which the responses were "spoiled" by the checking of more than one percept or rating were scored as "failure to make a selection."

To make sure that failure to make a selection did not represent a general negativism toward the test situation, the results of the subjects who had selected too few of the alternatives listed were eliminated.[2]

3. *Criterion Populations*

Scoring weights were determined from several criterion populations as follows:

a. *Set C1.* Weights were established upon this population for the purpose of scoring the same population. It was recognized that such treatment would lead to a spuriously high differentiation since all "chance" differences were weighted in favor of greater differentiation. However, if little differentiation resulted from this treatment, then cross validation would be unnecessary, as no better results could be expected.

b. *Sets C1 and C2 were combined to form population C12.* This was the criterion population upon which weights were developed for cross validation on Set X.

c. *Sets C1, C2 and C3 were combined forming population C123.* This was also used as a criterion population upon which weights were developed for cross validation on Set X. Population C123 was used to obtain information regarding the effect of:

(1). Increasing the size of the criterion population.

(2). Increasing the size of the critical ratio required for a response to be given a scoring weight. There were too few responses given weights in criterion population C12 when a p of .02 or .01 was used as a standard to permit an adequate judgment concerning these standards.

[2] Except for a few anxiety reactions in Set C1, subjects were required to select specific alternatives in at least 60% of the questions in each test, and in 75% of the test battery as a whole. In the total population of 653 subjects, only 6 subjects responded properly to less than 80% of the questions in the battery, only 13 responded properly to less than 85%, and only 29 responded properly to less than 90%. Furthermore, whenever it was felt that a subject had made his selections capriciously, as was suspected when the response checks formed an arithmetic or geometric pattern, no data from this subject were tabulated or scored.

4. *Weighting Systems Used with Different Criterion Populations*

a. With population C_1, only weighting system G was used.

b. With population C_{12}, the weighting systems used were:

(1). G
(2). $p = .10$
(3). $p = .05$

c. With population C_{123}, the weighting systems used were:

(1). G
(2). $p = .10$
(3). $p = .05$
(4). $p = .02$
(5). $p = .01$

5. *Cross Validation*

All cross validation was done on population X. The statistics used to evaluate the differences found between two distributions were:

a. *The percentage of subjects identified incorrectly at the best critical score.* Normal subjects scoring above the critical score were considered incorrectly identified, as were anxiety reactions with scores below it. The critical score was placed at that point which would result in the minimum percentage of subjects incorrectly identified in the total population; that is, the minimum number obtained when the number of anxiety reactions incorrectly identified was added to the number of normals incorrectly identified.

Two "best" critical scores were found empirically, one based on the distributions of the criterion population, and the other based on the distributions of the cross validation population. The percentage of subjects incorrectly identified in the cross validation population at each of these critical scores was determined. It is recognized that the exact percentage of subjects incorrectly identified at a critical score will not be highly reliable. However, an indication of the possible variability can be obtained by noting the percentage of subjects incorrectly identified at a point a few scores higher and a few scores lower than the critical score.

Furthermore, the most desirable critical score would vary greatly with the purpose for which a screening measure was used. As Ellis and Conrad (10) have pointed out, in most practical situations the percentage of false negatives or the percentage of false positives, as the case may be, is of greater importance than a combined percentage of wrong identifications. Thus, a much higher critical score would be employed when the purpose of the test was to definitely eliminate anxiety reactions than when the test was used to select possible anxiety reactions for a more thorough evaluation. However, the present study is attempting to determine generally whether any psychopathological screening is possible with a multiple choice projective test, and consequently it is concerned with the largest amount of differentiation possible, rather than with the direction of the error.

b. *The significance of the difference between the two distributions.* This was determined as a critical ratio.

V. RESULTS

A. Preliminary Analysis upon Population C_1

After population C_1 had been tested, a preliminary analysis of the data was made to determine whether further testing under the same experimental conditions was indicated. Scoring weights were developed from the responses made by the subjects in C_1 using the Guilford formula (weighting system G). Then these weights were used to score the subjects in C_1.

Since half of the "normal" subjects in C_1 took only Test AB, while the other half took only Test C, it was not possible to obtain a full battery score for these subjects. Therefore, the results of Test AB and Test C were treated separately.

A comparison of the Test AB scores received by the normals and the anxiety reactions revealed that only 9% were incorrectly identified at the best critical score. For Test C, only 6% were incorrectly identified. In the normal-schizophrenic comparisons, only 8% were incorrectly identified with each of the two tests. An anxiety reaction-schizophrenic comparison was made only for Test AB where 12.5% of the subjects were incorrectly identified. In all comparisons, the critical ratios were over 10. These results appeared definitely to warrant further testing with the SMCR.

An analysis was made of the results obtained with Tests A, B and C separately, and with various sections within a test, to determine whether any modifications should be made in the battery before further testing was undertaken. However, the differences in test conditions which were investigated, such as in the time of stimulus exposure, appeared to produce little difference in the results obtained so that no changes were made in the test battery.

B. Cross Validation of the SMCR

In the data presented below, unless otherwise specified, the results will always pertain to the subjects in population X; that is, the cross validation population.

1. *Discrimination Obtained with Different Weighting Systems*

a. *Percentage of subjects incorrectly identified.* Table 2 presents the scores obtained by normal and anxiety reactions subjects when different sets of weights are used.

In this table, as elsewhere, N represents the normal subjects and AR the anxiety reactions. Under "Score" are presented the lower scores in each step interval. To the right of each score are the number of subjects found in that step interval. At the bottom of the page, "M" "SD," and "CR" have their conventional meanings. The "% Wrong" stands for the percentage of subjects incorrectly identified at the best critical score.

Tables 3 and 4, presenting the results of the normal-schizophrenic and anxiety reaction-schizophrenic comparison pairs, are similar in form to Table 2.

The amount of discrimination obtained for the three comparison pairs with different sets of weights is summarized in Tables 5A and 5B which present the percentage of subjects incorrectly identified (% Wrong) and the critical ratios respectively.

As can be most easily observed from the summary tables, 5A and 5B, the various sets of weights produce a remark-

TABLE 2

Results Obtained When the Normal and Anxiety Reaction Cross-Validation Populations Are Scored with Different Sets of Weights

Criterion Score	Weights Based on C 12						Weights Based on C 123									
	Weighting System						Weighting System									
	G		.10		.05		G		.10		.05		.02		.01	
	N	AR	N	AR	N	AR	N	AR	N	AR	N	AR	N	AR	N	AR
20								2		4						
16	1						2	3		3	1	3				
12		3	1	2			1	1	3	2		4		1		
8		5	2	9	2	5	2	6	1	6	2	4	2	8	1	5
4	2	5		8	1	16	2	13	3	7	5	10	4	7	5	8
0	2	9	3	8	9	10	7	7	8	10	11	13	16	20	13	25
−1	7	6	14	11	21	13	13	8	10	5	11	7	22	9	30	10
−5	8	11	14	6	17	5	14	6	5	5	15	6	6	4	1	2
−9	14	4	11	5		1	7	2	11	5	5	2		1		
−13	8	4	4	1			2	2	9	2		1				
−17	4	2	1													
−21	3	1								1						
−25	1															
M*	−9.4	−2.3	−5.7	0.4	−2.4	1.3	−2.9	2.5	−4.8	2.4	−1.8	2.6	−0.5	2.0	−0.5	2.1
SD*	7.8	8.6	6.2	7.1	3.8	5.0	7.3	8.5	7.9	10.1	5.8	7.0	3.8	5.0	2.9	3.9
CR*	4.3		4.5		4.1		3.4		3.9		3.4		2.7		3.8	
% Wrong†	28		29		31		31		30		33		30		31	

* These calculations were made from data grouped with a class interval of 2. Because such a table was too large for one printed page, the class intervals were enlarged to 4 in the above table.

† These calculations were made from ungrouped data because in this statistic, unlike the others, grouping errors will occur in only one direction; that is, to increase the % Wrong.

TABLE 3

Results Obtained When the Normal and Schizophrenic Cross-Validation Populations Are Scored with Different Sets of Weights

Criterion Score	Weights Based on C 12						Weights Based on C 123									
	Weighting System						Weighting System									
	G		.10		.05		G		.10		.05		.02		.01	
	N	S	N	S	N	S	N	S	N	S	N	S	N	S	N	S
42								1								
36		1						4								
30		2								1						
24		1		1						1		2				
18		4		2			1	1		2		2		1		
12	1	1		1		1	1	1		3				3	1	2
6		2	1	6		3	2	8	1	1	1	4	1	3	1	3
0	3	10	1	3	2	6		11		7	1	9	1	9		15
−1	3	10	2	10		10	5	8	2	9	4	10	4	12	7	10
−7	2	5	4	10	6	14	2	7	4	7	3	6	2	10	6	12
−13	5	7	6	8	4	7	3	4	4	5	3	9	11	6	20	6
−19	9	3	6	3	16	8	10	3	5	7	11	4	20	5	11	1
−25	9	2	10	5	18	1	13	1	7	5	16	3	9		4	1
−31	9	2	12	1	3		9	1	8		6	1	2	1		
−37	7		6		1		2		12	1	4					
−43	1		1				2		5	1	1					
−49	1		1						1							
−55									1							
M*	−23.7	−2.2	−25.0	−7.2	−21.9	−8.1	−21.6	0.9	−29.7	−7.9	−23.3	−5.8	−18.5	−5.2	−14.4	−3.9
SD*	13.4	16.1	12.1	13.4	7.6	9.5	14.3	16.5	14.0	16.3	11.6	13.2	8.9	10.5	8.3	8.7
CR*	7.1		6.9		7.9		7.2		7.1		7.0		6.8		6.1	
% Wrong†	20		22		18		18		21		20		19		22	

* These calculations were made from data grouped with a class interval of 3. Because such a table was too large for one printed page, the class intervals were enlarged to 6 in the above table.

† The calculations were made from ungrouped data because in this statistic, unlike the others, grouping errors will occur in only one direction, that is, to increase the % Wrong.

TABLE 4

Results Obtained When the Anxiety Reaction and Schizophrenic Cross-Validation Populations Are Scored with Different Sets of Weights

Criterion	Weights Based on C 12						Weights Based on C 123									
	Weighting System						Weighting System									
Score	G		.10		.05		G		.10		.05		.02		.01	
	AR	S	AR	S	AR	S	AR	S	AR	S	AR	S	AR	S	AR	S
20		1						1								
16		3						3		1						
12		3		3				1		1						
8	2	5	1	5		1	1	5				4		1		
4	1	4	1	5		5	3	5	3	7		3		3	1	5
0	4	3	3	4	2	8	1	2	1	3	3	6	6	14	4	17
−1	2	8	5	10	7	11	7	12	4	8	4	10	4	13	15	15
−5	2	9	2	9	9	12	3	7		7	6	10	14	11	17	10
−9	7	5	8	6	6	5	8	6	5	9	9	9	18	4	13	3
−13	6	3	12	2	17	3	6	2	8	7	11	2	8	4		
−17	6	1	8	2	9	2	11	2	7	1	13	5				
−21	10	3	6	3		1	8	3	9	1	4	1				
−25	3	1	4				2	1	7	4						
−29	5			1					5							
−33	2	1							1	1						
N*	−15.4	−2.7	−13.2	−3.8	−11.0	−4.6	−12.3	−2.6	−17.3	−7.4	−12.7	−5.3	−8.1	−2.9	−5.7	−1.7
SD*	11.2	12.6	8.9	10.7	6.0	8.3	8.9	11.1	10.1	10.4	6.4	7.8	4.6	5.5	3.9	3.9
CR*	5.3		4.7		4.3		4.8		4.8		5.2		5.1		5.1	
% Wrong†	24		24		27		27		24		26		25		31	

* These calculations were made from data grouped with a class interval of 2. Because such a table was too large for one printed page, the class intervals were enlarged to 4 in the above table.

† These calculations were made from ungrouped data because in this statistic, unlike the others, grouping errors will occur in only one direction, that is, to increase the % Wrong.

TABLE 5A

Summary of Percentage of Subjects Incorrectly Identified (% Wrong) in the Cross-Validation Sample

Comp. Pair	Weights Based on Population C 12			Weights Based on Population C 123					Mean C 12	Mean C 123
	G	.10	.05	G	.10	.05	.02	.01		
N-AR	28	29	31	31	30	33	30	31	29	31
N-S	20	22	18	18	21	20	19	22	20	20
AR-S	24	24	27	27	24	26	25	31	25	27
Mean	24	25	25	25	25	26	25	28	25	26

TABLE 5B

Critical Ratios for Statistics in Table 5A

Comp. Pair	Population C 12			Population C 123					Mean C 12	Mean C 123
	G	.10	.05	G	.10	.05	.02	.01		
N-AR	4.3	4.5	4.1	3.4	3.9	3.4	2.7	3.8	4.3	3.4
N-S	7.1	6.9	7.9	7.2	7.1	7.0	6.8	6.1	7.3	6.8
AR-S	5.3	4.7	4.3	4.8	4.8	5.2	5.1	5.1	4.8	5.0
Mean	5.6	5.4	5.4	5.1	5.3	5.2	4.9	5.0	5.5	5.1

ably similar amount of discrimination. For each comparison pair, the middle six (of eight) "% Wrong" figures have a range of only three percentage points, while the critical ratio range is less than one point. This consistency clearly indicates that differences both in the weighting systems and in the criterion populations upon which the weights are based have had little effect upon the amount of discrimination produced.

b. *Particular subjects incorrectly identified.* It is possible that while the same percentage of subjects was incorrectly identified by each set of weights, there may be differences with respect to the particular subjects who were incorrectly identified. Therefore, a comparison of the subjects incorrectly identified by each set of weights was made.

For each diagnostic group in a comparison pair, there were eight distributions, a different one produced by each set of weights. The 10 subjects scoring lowest or highest, whichever indicated incorrect identification, were selected from each distribution as the ones most clearly misidentified. Then, for each diagnostic group in a comparison pair, the 10 subjects appearing in most distributions as incorrectly identified were considered the criterion "incorrectly identified groups." It was found that there was very little difference between the 10 subjects selected by any of the sets of weights and the subjects in the criterion group.

Because of the marked similarity in the results obtained with the various sets of weights, it was felt justifiable to use only one set in the subsequent treatment of the data. The weighting system $p = .10$ with weights derived from population C_{123} was selected because it identified so many of the subjects in the incorrectly identified group and because it led to the widest range in scores, an important factor when the results obtained from only sections of the test battery were considered.

2. *Effect of Increasing the Number of Subjects*

The fact that adding the subjects in population C_3 to the criterion population did not increase the validity may be attributed to one of two reasons. Either the 100 subjects in population C_{12} provided a sufficiently large number to attain the best possible set of weights, or the poorer diagnostic classification of the subjects in C_3 was enough to balance any additional validity to be gained from adding 50 subjects to each category in the criterion group. To investigate this problem, populations C_2 and C_3 were scored separately with weighting system $p = .10$ based on criterion population C_{123}. Population C_1 was not scored because a total score could not be obtained for the "normal" subjects since each of them had taken only half of the testing battery.

In the N-AR comparison pair, the percentage of subjects incorrectly identified was 18% for C_2 and 16% for C_3. In the N-S comparison pair, these percentages were 16% and 17% respectively, while for the AR-S comparison pair they were 19% and 25% respectively. While it is true that C_2 and C_3 each formed one-third of the criterion population upon which the weights were based, still if the diagnostic classification of the subjects in C_3 was materially inferior, this population should have shown much less discrimination. Since the percentage of subjects incorrectly identified was very similar for both C_2 and C_3, it would seem more logical to conclude that for

this test the diagnostic classification of subjects in C_3 could not be much inferior to that of the subjects in C_2. Hence, for the SMCR, the addition of subjects to the criterion population beyond 100 per diagnostic category would not seem likely to materially increase the amount of discrimination produced.

3. *Reliability of the SMCR*

Because subjects were not available for retesting, the reliability was determined by a measure of internal consistency. The test battery was divided into four sections: Test A, Test B, the first half of Test C, and the second half of Test C. Each section was then split into two halves according to a table of random numbers. Scores on one half of each section were summed to give a Part 1 score which was correlated against a Part 2 score derived from the sum of scores on the other half of each section.

For each of the three comparison pairs, reliabilities were found for each diagnostic group separately, and then for both groups combined. In the N-AR comparison, the reliability coefficients, when corrected by the Spearman-Brown prophecy formula, were $.69 \pm .07$ for the normals alone, $.75 \pm .06$ for the anxiety reactions alone, and $.77 \pm .04$ when both groups were combined. In the N-S comparison, the corrected coefficients were $.78 \pm .05$ (N), $.82 \pm .05$ (S), and $.87 \pm .02$ (N-S); while those for the AR-S comparison were $.65 \pm .08$ (AR), $.67 \pm .08$ (S), and $.72 \pm .05$ (AR-S). These results would indicate that the SMCR has a reasonable degree of reliability.

4. *Validity of the SMCR*

There would seem to be no question that valid group differentiations have been made for each comparison pair. In Table 5B only one of the critical ratios was under 3.0, and that came to 2.7. However, the amount of error when individual prediction is desired can be gathered only from the percentage of subjects incorrectly identified. In Table 5A, when weighting systems $p = .01$ was eliminated because it was the only one which gave somewhat atypical results, the median of the other seven weighting sets indicate that 30% of the subjects in the N-AR comparison pair are incorrectly identified, as are 20% in the N-S comparison pair and 25% in the AR-S comparison pair.

However, it must be recognized that these percentages are based on the "best" critical score of the cross validation distributions. This procedure weights all sampling errors in favor of a greater discrimination, and thus gives a spuriously high validity. To properly estimate validity, the critical score should be based on a previous population. Therefore, the best critical score was found from the distributions of population C_{23} scored with weighting system $p = .10$ based on criterion population C_{123}. These critical scores proved to be $+2$ for the N-AR comparison pair, -21 for the N-S comparison pair, and either -7 or -8 for the AR-S comparison pair. There were two possible critical scores for the last pair because both gave the same percentage of subjects incorrectly identified.

At the above cutting scores there would be 30% of the N-AR and 25% of the N-S comparison pairs incorrectly identified. Since a cutting score of -8 would result in 31% of the subjects being incorrectly identified, while a score of -7 would result in 29% of the subjects being incorrectly identified, it would ap-

pear best to estimate the percentage of subjects incorrectly identified in the AR-S comparison pair at 30%.

It should be recognized that using the results of only one rather than a series of samples to establish the critical scores, and having too few subjects in the distributions for smooth curves may lead to an underestimate of the validity. Therefore, the best possible estimate of the validity may lie somewhere between the above results and those obtained when using the best critical score of the cross validation distribution. However, exactly where in between these points it would lie is impossible to determine, and it would appear safest to conclude that the SMCR will incorrectly identify at most 30% of the N-AR comparison pair, 25% of the N-S comparison pair, and 30% of the AR-S comparison pair.

5. *Classifying Subjects According to Tri-Scores*

In the above estimate of validity, the subjects in a comparison pair were scored only with the set of weights derived from the distributions of that comparison pair. However, these subjects could also be scored with the weights derived from the other two comparison pairs. Thus, it was possible to score each subject with three sets of weights, one from each comparison pair, and thereby classify him:

1. As normal (N) or as anxiety reaction (AR) from the results obtained with the N-AR comparison pair scoring key.
2. As N or S from the N—S comparison pair scoring key.
3. As AR or S from the AR-S comparison pair scoring key.

Each subject could then be classified in one of the following eight ways: (These classification combinations will be referred to in the future as tri-scores.)

1. As N on the N-AR key; N on the N-S key; and AR on the AR-S key. NNA
2. N (N-AR key); N (N-S key); S (AR-S key). NNS
3. N (N-AR key); S (N-S key); AR (AR-S key). NSA
4. N (N-AR key); S (N-S key); S (AR-S key). NSS
5. AR (N-AR key); N (N-S key); AR (AR-S key). ANA
6. AR (N-AR key); N (N-S key); S (AR-S key). ANS
7. AR (N-AR key); S (N-S key); AR (AR-S key). ASA
8. AR (N-AR key); S (N-S key); S (AR-S key). ASS

To determine the validity obtained by using these tri-scores, the same validation procedure was followed as when subjects in a comparison pair were scored only with the set of weights derived from that comparison pair.

Population C_2 was used as the criterion population, and the tri-scores of this group were found. The diagnostic identification of these subjects was based on whether they scored above or below a critical score. For each comparison pair the critical score was determined from the distributions of the two diagnostic groups within that comparison pair. (In all comparison pairs, the point at which the critical score fell proved to be the same for population C_2 as had been found for population C_{23}.)

For the cross validation population, two separate tri-score distributions were obtained; one where the diagnostic classification of the subjects was based on a critical score derived from the distribution of the cross validation population itself (Pop. X), and the other where the diagnostic classification was based on the point selected as the critical score in population C_2. (Pop. X′). The tri-score distributions are presented in Table 6.

The tri-score weights were determined for each comparison pair by the tri-score

TABLE 6

NUMBER OF SUBJECTS IN EACH TRI-SCORE CLASSIFICATION

Tri-Score	Population C_2 (Classified according to own "best" critical score)			Population X (Classified according to own "best" critical score)			Population X′ (Classified according to C_2 critical score)		
	N	AR	S	N	AR	S	N	AR	S
NNA	39	12	5	29	14	3	32	19	8
NNS	0	1	0	1	1	0	1	0	1
NSA	1	1	2	4	5	4	6	1	6
NSS	6	0	14	6	0	20	1	0	12
ANA	1	8	0	2	6	0	4	9	3
ANS	0	0	1	0	1	0	0	1	0
ASA	0	20	3	3	11	3	3	12	7
ASS	3	8	25	5	12	20	3	8	13

distributions of the two diagnostic groups in a comparison pair. For example, in the N-AR comparison pair, the tri-score NNA was considered as a normal score since the population C_2 section of Table 8 reveals that 39 normals and only 12 anxiety reactions made this tri-score. For two tri-scores, NSA and ANS, there were the same number of normals and anxiety reactions making these scores. In such cases, the tri-score was weighted according to the results of the comparison pair involving the two groups concerned. Thus, in the N-AR comparison, the results obtained with the N-AR set of weights would determine whether a subject would be classified as normal or as anxiety reaction.

It can be seen in Table 7 that for both group X and X′, the % Wrong using only the critical score in a subject's comparison pair gives about the same results as the % Wrong using the tri-score. The very limited improvement which results when the best possible combination of tri-scores is found for each group indicates that little could be expected from using a criterion population larger than C_2 in establishing the tri-score weights.

However, the tri-score is not entirely without advantage. Thus, for subjects receiving the tri-score NNA, a statement could be made that they are most likely to be normals; but that if they are not normals, then they are more likely to be anxiety reactions rather than schizophrenics. There are a sufficient number of subjects with tri-scores NNS, ANA, ASA, and ASS to make similar interpretations possible for subjects in these clas-

TABLE 7

PERCENTAGE OF WRONG DISCRIMINATIONS OBTAINED WITH CRITICAL SCORES ONLY, AND WITH TRI-SCORES

	Population C_2			Population X			Population X′		
	N	AR	S	N	AR	S	N	AR	S
With critical score only	18	16	19	30	21	24	30	25	31
At best tri-score combination	17	15	17	29	20	22	30	25	27
At tri-score combination using C_2 tri-score wts.				30	21	24	31	25	29

sifications. Of the 134 subjects in population X′ receiving these tri-scores, a prediction as to which of the three diagnostic groups a person is *not* in would be incorrect in only 17, or 13%, of the cases. Had the diagnostic prediction of the 16 subjects not receiving any of the above tri-scores been made on a chance basis, the percentage incorrectly identified would still be under 17%.

6. *Comparison of A Priori Judgments as to Diagnostic Significance of Responses with Empirical Results*

The a priori judgments were compared to the empirical results in regard to the diagnostic implications (normal, anxiety reaction or schizophrenic) of the various percepts. The results of this comparison are presented in Table 8. Listed at the left are the diagnostic categorizations which were determined empirically. These categorizations were based on the results obtained with the entire experimental population. The letter "O" indicates those percepts which did not receive any weights (using the $p = .10$ weighting system), and hence could not be considered characteristic of either diagnostic group.

The a priori judgments as to diagnostic implication are presented along the top of the table. The "?" stands for the alternative "something else," while the letter "F" is the response "failure to select one of the alternatives in a question." Separate results were computed for Test AB and Test C. In the former, it was possible to determine the empirical weight for each percept exactly. In the latter, however, this determination had to be made qualitatively.

In Test C, the subjects checked one of five possible ratings for each percept, and a separate weight was found for each of these five ratings. For some percepts, the diagnostic implication was quite clear since the first few rating steps were characteristic of one diagnostic category, while the last few were characteristic of the other diagnostic group. However, the picture generally was more confused. For example, the first and fifth steps might be given a normal weight, the fourth step an anxiety reaction weight, and the second and third steps given no weight. With results such as these, it was impossible to define the percept as characteristic of normals or of anxiety reactions. In many cases, borderline decisions had to be made, such as when of the five rating steps only the first or second step was found to be weighted. Included in the present table were only those percepts whose diagnostic implications seemed fairly clear.

Table 8 reveals that for the N-AR comparison pair there is no correlation between a priori judgment and the empirical results.[1]

Of the 60 percepts considered to be characteristic of normals on Test AB, only 9 proved to be so empirically, and in four cases the supposed normal percepts were found to be characteristic of anxiety reactions. Of the 80 percepts judged to be characteristic of anxiety reactions, only 6 were so empirically, while 13 were found to be characteristic of normals. Thus, in only 15 of the 140 percepts did the a priori judgment coincide with the empirical evaluation, while in 17 cases they were exactly the op-

[1] It should be clearly recognized that the authorities who estimated the psychopathological significance of the responses did so on the basis of their experience with the individual Rorschach, and they made no claims that the significance would remain the same on a multiple choice test.

posite. Very similar results were also obtained on Test C.

For the other two comparison pairs, there appears to be a positive correlation between a priori judgment and empirical evaluation. In the N-S comparison, of the 60 percepts presumed to be normal on Test AB, 16 actually were so with only 5 being exactly the opposite; that is, being characteristic of schizophrenics. 29 percepts coincide, and 25 are exactly the opposite.

It is difficult to use a single statistic to indicate the amount of correspondence because a lack of coincidence between the a priori judgment and the empirical results may be expressed in two ways: the estimated percept may prove to be significant of neither diagnostic category; or, what appears to be more serious, it

TABLE 8

CORRESPONDENCE BETWEEN EMPIRICAL AND A PRIORI CATEGORIZATIONS

Comparison Pair	Empirical Categorization	A Priori Categorization								
		Test AB					Test C			
		N	A	S	?	F	N	A	S	F
N-AR	Normal	9	13	6	2	0	14	12	2	0
	Anxiety Reaction	4	6	19	4	1	4	5	16	5
	Neither	47	61	55	14	59				
N-S	Normal	16	19	6	4	0	17	8	6	0
	Schizophrenic	5	12	29	5	35	13	9	14	67
	Neither	39	49	45	11	25				
AR-S	Anxiety Reaction	14	15	10	2	0	18	19	1	0
	Schizophrenic	3	15	14	1	26	12	14	23	45
	Neither	43	50	56	17	34				

Of the 80 percepts estimated as schizophrenic, 29 were found to be so empirically with only 6 having exactly the opposite diagnostic significance. Thus, in 45 of the 140 percepts the a priori judgments and the empirical results coincided, while they were exactly the opposite for only 11 percepts. The results of Test C, though, are not quite as good, with only 31 of the percepts coinciding while 19 were exactly the opposite.

In the AR-S comparison it is Test C which shows the greater amount of correspondence between a priori judgment and empirical evaluation. Here 42 of the percepts coincide, with only 15 being exactly the opposite. However, little relationship is shown in Test AB where may turn out to be significant of exactly the opposite diagnostic category than was expected.

In general, it may be said that a positive correlation exists between a priori judgments and empirical results since these two coincide more frequently, on the whole, than they are exactly the opposite. However, this correlation seems to be fairly low as can be seen from the relatively few times a percept is found empirically to have exactly the same significance as was expected on an a priori basis. Thus, while a priori judgments may be of value in the initial selection of items for a multiple choice projective test, there can be no question but that it will be essential that the

psychopathological significance of each percept be determined empirically.

However, the data in Table 8 also indicate that a detailed analysis of those percepts found to be significant of a diagnostic category may uncover general principles applying to multiple choice projective tests which could materially increase the validity of a priori judgments. For example, in the N-AR comparison, the a priori estimations would have been much more accurate had those percepts which were considered to be characteristic of schizophrenics been judged to be indicative of anxiety reactions, while those percepts predicted as typical of anxiety reactions were instead considered characteristic of normals. This holds true for both Test AB and Test C. Similarly, in the AR-S comparison, the percepts which were thought to be indicative of anxiety reactions proved to be much less discriminating than the percepts which has been assumed to be characteristic of normals. Also apparent from Table 8 is the fact that the alternative "Something else" (?) does not appear to differentiate any diagnostic group, while the "Failure to select an alternative" (F) seems to be primarily a schizophrenit response.

7. *Comparison of Conditions Varied within the Test Battery*

As outlined previously (Section II,B, 2), tne experimental design included variations in certain of the test conditions. By comparing the results of different sections of the test battery, the effect of these variations could be determined. The relative discrimination obtained under the different test conditions is shown in Table 9. While the amount of discrimination under each condition was measured by both the critical ratio and the % Wrong (percentage of subjects incorrectly identified), only the critical ratios are listed because of the extremely high correlation between these two statistics. Preference was given to the critical ratio because it would seem to be the more reliable statistic.

a. *Differences in the structure of the questions.*

(1). *Relative vs. absolute decision by the subject.* In Test A and Test B,

TABLE 9

COMPARISON OF CONDITIONS VARIED WITHIN THE TEST BATTERY

Condition	Test Section	Critical Ratios for Different Comparison Pairs		
		N-AR	N-S	AR-S
Relative vs. absolute decision by subject	Test AB (relative)	3.4	5.5	4.0
	Test C (absolute)	3.0	7.8	4.1
Difference in type of percepts in a question	Design I (forced normal)	0.7	4.1	2.7
	Design II (unlimited)	3.8	3.7	2.8
	Design III (forced non-normal)	1.6	4.6	3.0
Brief vs. lengthy exposure of stimulus	Test A (brief)	2.6	4.1	3.5
	Test B (lengthy)	3.2	5.9	2.5
No. of questions in the test battery	Designs I, II, III (20 questions)	2.0	4.5	2.8
	Test A, B (30 questions)	2.9	5.0	3.0
	Test AB (60 questions)	3.4	5.5	4.0
	Test C (100 questions)	3.0	7.8	4.1
	Test ABC (160 questions)	3.9	7.1	4.8

which have been combined to form Test AB, the subjects were required to select one of four alternatives, thus making a relative decision among these alternatives. On Test C, the subjects rated each alternative on a five point absolute rating scale. As can be seen from the results in Table 9, the type of decision made by the subjects appears to have little effect upon the amount of discrimination. While Test AB is somewhat superior for the N-AR comparison pair, Test C has a higher critical ratio for the N-S comparison pair, while the results are practically identical in the AR-S comparison pair.

(2). *Differences in the types of alternatives included in a question.* Results were obtained for three different question "Designs," each of which was structured to include only certain types of percepts (section II,B,2a [2]). It was necessary to select percepts on the basis of an a priori judgment as to their psychopathological significance, and, as has been shown, such a priori judgments do not correspond to the psychopathological significance found empirically. Hence, the original rationale for this comparison no longer exists. However, there remains a comparison between questions in which the subjects are required to select a specific percept (Designs I and III), and questions in which the subject may avoid selecting a specific percept by checking the alternative, "Something else," (Design II).

From the results it is seen that Design II was superior for the N-AR comparison pair, while Design III was best for the N-S comparison pair. Little difference was found among the designs in the AR-S comparison pair. While no superiority seems to be shown for either the "forced-choice" or the "unlimited-choice" question designs, any conclusions should be made cautiously since these results were based on only 20 questions of each design.

b. *Differences between a brief and lengthy exposure of the stimulus.* In Test A the stimuli were exposed for a relatively brief period. In Test B, which was equivalent to Test A in other respects, the subjects were given as much time as they desired to view the stimulus. As shown in Table 9, Test B produced greater discrimination in the N-AR and N-S comparison pairs, while Test A was superior in the AR-S comparison pair. These results do not appear to be definitive enough to warrant drawing any general conclusions. It remains possible that the differences found represent true differences, and that opposite results are obtained in different comparison pairs because "optimum" conditions are different for each pair of diagnostic groups.

c. *The effect of increasing the length of the test.* By considering different sections of the test battery separately, it was possible to determine empirically the effect of increasing the length of the test. Designs I, II and III each included only 20 questions, so the mean of the critical ratios obtained with these "Designs" was used to estimate the discrimination produced by 20 questions. The amount of discrimination resulting from 30 questions was determined by the mean of the critical ratios obtained with Test A and Test B, each of which contained that number of questions. When Tests A and B were combined to form Test AB, there was a total of 60 questions. In Test C there were 100 questions, while the complete battery, Test ABC, contained 160 questions. The fact that the types of questions were different for different test

lengths was not considered contaminating since it would appear from the preceding results that the type of question has little effect upon the amount of discrimination.

As can be seen in Table 9, there is, in general, a steady increase in discrimination for each comparison pair as the Wrong of the three comparison pairs drops from 35% for the 20 question section to 25% for the 160 question battery.

8. *The Effect of Age and Education*

Since there were some differences among the three diagnostic groups with respect to age and amount of education,

TABLE 10

COMPARISON OF SUBJECTS INCORRECTLY IDENTIFIED AND SUBJECTS CORRECTLY IDENTIFIED, WITH RESPECT TO AGE AND AMOUNT OF EDUCATION

Diagnostic Category	Compar. Pair	Group	Age Comparison			Education Comparison		
			N	Mn	Mdn	N*	Mn	Mdn
Normals	N-AR	Incorrect	10	25.2	25.2	10	11.7	12.5
		Correct	40	25.8	24.3	40	12.5	12.7
	N-S	Incorrect	10	25.5	23.5	10	11.6	11.7
		Correct	40	25.7	25.1	40	12.5	12.8
	Both	Pop. X	50	25.6	24.7	50	12.3	12.7
		Total Pop.	250	25.0	23.8	245	13.2	12.8
Anxiety Reactions	N-AR	Incorrect	10	33.3	33.0	10	12.3	13.0
		Correct	40	30.1	29.5	40	10.6	10.6
	AR-S	Incorrect	10	35.1	36.0	10	8.8	8.8
		Correct	40	29.6	29.5	40	11.4	12.1
	Both	Pop. X	50	30.7	30.5	50	10.9	10.9
		Total Pop.	203	30.0	29.6	197	11.4	11.5
Schizophrenics	N-S	Incorrect	10	27.3	25.5	8	12.3	12.0
		Correct	40	29.2	28.5	34	11.0	11.6
	AR-S	Incorrect	10	26.6	25.5	9	11.8	11.5
		Correct	40	29.4	28.8	33	11.1	11.6
	Both	Pop. X	50	28.8	28.0	42	11.3	11.6
		Total Pop.	200	28.4	27.3	179	11.5	11.6

* For some subjects information as to the amount of education was not available.

length of the test increases. Were the critical ratios of the three comparison pairs to be averaged, then the few fluctuations would be eliminated, and the amount of discrimination would show a regular progression. The importance of the length of the test upon individual prediction can be gathered from a comparison of the percentage of subjects incorrectly identified. The mean % it was possible that the discrimination obtained between diagnostic groups was a reflection, in part at least, of these age and educational differences. In Table 10, the correctly and incorrectly identified subjects (defined in section V,B,1,b) in each comparison pair are differentiated with respect to age and amount of education. For purposes of comparison, data are also presented for all of popula-

tion X, which includes both correctly and incorrectly identified subjects, and for the entire experimental population.

Each diagnostic group is included in two comparison pairs. The results of the normals and schizophrenics are very similar in both comparison pairs because in each case many of the same subjects were incorrectly identified in both comparison pairs. Six of the normals who were incorrectly identified in the N-AR comparison pair were also incorrectly identified in the N-S comparison pair. With the schizophrenics, there were nine subjects duplicated. However, none of the anxiety reactions was incorrectly identified in both comparison pairs. (A chance matching of 10 names in each diagnostic category would have resulted in two subjects being included in both comparison pairs.)

There seems to be no difference in age between correctly and incorrectly identified normals. However, the incorrectly identified schizophrenics are a few years younger than those correctly identified, while the incorrectly identified anxiety reactions, particularly in the AR-S comparison pair, are an older group.

The incorrectly identified normals have slightly less education, while the incorrectly identified schizophrenics are slightly better educated. With the anxiety reactions, however, the educational differences are marked. Those incorrectly identified as normals have about two more years of education than those correctly identified, while the anxiety reactions identified as schizophrenics have about three years less education.

From these results it would appear that, particularly in the case of the anxiety reactions, the subjects' responses are a product of their educational background as well as of their psychopathological condition. The educational background may, in turn, be a reflection of intellectual endowment. The entire question appears to be too complex to be resolved in the present study, particularly since the inadequacy of the original diagnostic classification may also be a factor. The chief screening psychiatrist, in charge of the panel which diagnosed almost all the anxiety reactions in the cross validation population, stated that most diagnostic errors probably occurred when attempting to determine whether the older patients with little education (i.e., those anxiety reactions identified on the SMCR as schizophrenics) were anxiety reactions or schizophrenics. In any case, the overlap between correctly and incorrectly identified subjects with respect to age and education was so great that differences in these factors cannot be responsible for most of the discrimination obtained with the SMCR.

VI. DISCUSSION OF THE RESULTS

Upon cross validation, the SMCR incorrectly identified at most 30% of the N-AR comparison pair, 25% of the N-S comparison pair, and 30% of the AR-S comparison pair. For all comparison pairs the critical ratios were sufficiently high to establish a significant difference between diagnostic groups. By using tri-scores, the diagnostic category a subject was *not* in could be predicted with an error of under 17%. Therefore, it would appear that a fair validity had been obtained with a multiple choice projective test. Since the amount of discrimination found compares favorably with the validity reported for other screening measures, and since there are indications that subsequent research with multiple choice projective tests will lead to even greater discrimination, further investigation with this approach would seem warranted.

A. Comparison with Other Screening Devices

As mentioned previously, screening tests have been found to be most effective in military screening. Ellis and Conrad (10) have presented a detailed summary of the results obtained with personality inventories in this area. For most studies, the amount of validity reported is approximately the same as that obtained with the SMCR. While there were a few reports of unusually high validity, Ellis and Conrad suggest that such validity coefficients reflect the operation of unusual experimental or statistical factors.

The possible effect of improper cross validation, an experimental error found in many studies, has been well illustrated by data obtained in the present investigation. When a population of 50 subjects in each diagnostic group (C_1) was scored with weights derived from the same population, the scores from just half the test battery resulted in only about 8% of the subjects being incorrectly identified. When weights derived from a criterion population of 150 subjects in each diagnostic category (C_{123}) were used to score one-third of these subjects (C_2 and C_3), the percentage of subjects incorrectly identified rose to an average of 18.5%. When the weights derived from this criterion population were applied to a cross validation group, the percentage of subjects incorrectly identified at the best critical score of the cross validation group became 25%. When a critical score derived from the criterion population was employed, the average percentage of subjects incorrectly identified rose to 28%.

Zubin (60), in another article evaluating screening tests, presents the results obtained at the Sheepshead Bay U. S. Maritime Training Station as an example of the marked effectiveness of a questionnaire in military screening which generally has produced better results than civilian screening. Of 1,000 men processed, 660 were passed without a further interview on the basis of a screening personality questionnaire. Three of those passed had to be discharged from the service during their training period for psychiatric reasons. Of the 340 men selected for a psychiatric evaluation, 15 had to be discharged from the service, 75 had temporary or mild emotional conditions which required short term therapy, while the remaining 250 turned out to be false positives.

In evaluating the above data, the 75 subjects who adjusted after short term therapy are difficult to place. There may have been an equal

proportion of subjects among the persons passed without a psychiatric interview who had comparably mild emotional conditions, but who were able to adjust without any short term therapy. If these 75 subjects were therefore to be considered as false positives, there would be a total of 1,000 subjects, 982 passing and 18 being discharged. Of these, 33% (325 passing and 3 discharged) were incorrectly identified. Eliminating the 75 subjects from the data would lower the percentage of subjects incorrectly identified to 27%.

Not only are the above percentages about the same as those obtained with the SMCR, but they probably give a spuriously high estimate of validity. Only persons making high questionnaire scores would have been discharged (on the basis of a psychiatric examination) without participating in the regular training program. Ellis and Conrad report a study by Weinstock and Watson "on 121 Naval recruits who were allowed to remain in service ('on trial') despite an adverse prognosis based on clinical judgment. Of the 121 recruits, only 44 (36%) were discharged for neuropsychiatric reasons during recruit training" (10, p. 386).

Another factor which frequently leads to spuriously high validity estimates is the use of extreme groups. With such populations almost any screening measure can obtain a fair validity. As Huston and Shakow found in a study on schizophrenics, while the patient groups performed at a level significantly poorer than the control group on two indices of steadiness, "when these patients who cooperated as well as the normal group were compared with the latter, the differences between the groups disappeared" (25, p. 120). It was for this reason that in the present study care was taken to test only the "best" schizophrenics, and to use anxiety reactions who were not so severely disturbed as to warrant hospitalization.

Had a typical institutionalized population been used, the SMCR would undoubtedly have produced a much greater amount of discrimination. For example, there were 48 schizophrenic subjects who filled out Tests A and B correctly according to the established criteria, but who failed to fill out Test C properly. The greater failure on Test C resulted from the fact that it required a somewhat more complex response and came last in the testing battery when increased boredom and fatigue had set in. When the scores of the 48 schizophrenics who failed to complete Test C were substituted for the regular schizophrenic cross validation population in the Test AB comparisons, the percentage of subjects incorrectly identified at the best critical score dropped from 24% to 10% in the N-S comparison, and from 32% to 22% in the AR-S comparison. Furthermore, had the SMCR been administered so that a higher proportion of schizophrenics would have responded, such as under individual conditions or with proctoral assistance, the percentage of subjects incorrectly identified would undoubtedly have been even smaller.

B. Indications for Greater Potential Discrimination

While there has been a tremendous amount of experimentation with questionnaires extending back to the first World War, relatively little work has been done with multiple choice projective tests. Indications exist from the present study that further research with the multiple choice projective approach may lead to a marked increase in the amount of discrimination produced.

Firstly, the present study did not attempt to exhaustively sift questions until a test composed of the most discriminatory choices was evolved. By far the greater number of responses received no weights, even under the most liberal weighting system. Yet the greater amount of discrimination occurring as the size of the test was increased clearly indicates

that the replacement of non-discriminatory percepts should add even more to the validity.

Secondly, only one projective method, the Rorschach, was used. It is recognized in individual psychopathological testing that much more validity can be obtained from a battery of different projective tests than from any single instrument. As Rapaport and Schafer have stated: ". . . as an exclusive basis of diagnosis the test (the Rorschach) has serious limitations and its real potency is exploited only when used as the center of a battery of tests" (42, p. 73).

Thirdly, a loss in validity undoubtedly occurred because of inadequate diagnoses of subjects in the three diagnostic classifications. The Cornell Index, which was used as the criterion of normality is itself only a rough screening device, and it has not proved too effective in the screening of prepsychotic and early psychotic states. Norms for Form N1 of the Cornell Index could not be obtained. However, 10 or fewer poor responses on Form N2 which contains 37 additional questions and consequently provides more opportunity for poor responses, will be made by 14% of the persons rejected for military service for psychiatric reasons. Thus, it is probable that some subjects in the "normal" population were poorly adjusted. Furthermore, some of the schizophrenics were in remission, or at least well enough adjusted to be released from the hospital within a few weeks of their testing, while others had a great deal of free floating anxiety characteristic of anxiety reactions. The anxiety reaction group was perhaps the most poorly diagnosed, containing many individuals who would not have contemplated psychiatric treatment had they not been veterans, as well as persons whose anxiety was symptomatic of a latent schizophrenia. Almost all the anxiety reactions in the cross validation population were selected only on the basis of a screening or tentative psychiatric diagnosis.

Of course, most subjects in each diagnostic category were probably correctly classified. However, the number of subjects incorrectly identified by the test was relatively small, so that the inclusion of some misclassified persons in this group would make an appreciable difference. It is recognized that the misclassified subjects may not be the persons incorrectly identified by the test. However, the possibility appears to be sufficiently great to warrant a more thorough evaluation of the subjects in each diagnostic category before the maximum potentiality of the test can be estimated.[1]

[1] Since a year had elapsed between the testing of the subjects and the preparation of this monograph for publication, an attempt was made to check the accuracy of the diagnosis established at that time by a follow-up study. No further information could be obtained for the normal subjects since they had filled out the tests anonymously. However, psychiatric progress reports were available for most of the anxiety reactions and schizophrenics.

The majority of the anxiety reactions tested were subsequently treated by psychiatrists working in private clinics under contract to the Veterans Administration. These psychiatrists prepared monthly reports summarizing the condition of their patients. The majority of the schizophrenics had been tested at the Veterans Hospital, Northport, L. I., and for these subjects complete clinical case folders were available.

Unfortunately, a careful inspection of this data was not helpful in improving the validity of the original diagnosis. While almost all of the subjects retained their initial diagnosis, a consideration of their symptomatology revealed much overlapping between subjects diagnosed as anxiety reactions and those diagnosed as schizophrenics. It appeared that psychiatrists at out-patient clinics generally preferred to make a diagnosis of anxiety reaction while those at neuropsychiatric hospitals were more likely to make a diagnosis of schizophrenia. This is in agreement with the experience of the writer at both types of installations, and with the opinions of several psychiatrists consulted on this question. Thus, subjects with identical symptoma-

Finally, a multiple choice projective test combined with other types of tests in a battery may result in a greater amount of discrimination than would exist for any of the tests separately. For example, a test such as the SMCR which resulted in greater discrimination between normals and psychotics might be combined with a questionnaire such as the Cornell Index which is more effective in differentiating normals from neurotics. Subjects might also be given an intelligence test, and their multiple choice projective test responses interpreted according to their I.Q. It is recognized that responses on the free association Rorschach will depend upon a subject's intelligence as well as his psychopathological condition. A "vague" or "mediocre" response which might be considered normal for a subject of low intelligence would be indicative of maladjustment in one of high intelligence. The same would appear to be true of the multiple choice Rorschach. In the present investigation, the large difference in education between the anxiety reactions identified incorrectly as normals and those identified incorrectly as schizophrenics probably reflects a difference in intelligence rather than in education per se. Therefore, it might be advisable to develop different sets of scoring weights for different levels of intelligence.

C. Suggestions as to Future Research

From the results of the present study, both quantitative and qualitative, a number of suggestions arise as to the course of future investigations.

1. *Subjects*

In the present study, weights based on the results of only 50 subjects in each diagnostic category proved inadequate. On the other hand, a criterion population of 150 subjects per diagnostic category was no more effective than one of 100 subjects. But caution would be necessary in applying these results to other investigations. For example, a smaller number of subjects might be required when only two or three responses are possible for each question rather than the 5 or 6 in the SMCR, or when a psychopathological variable less heterogeneous than diagnostic categories is investigated. However, the lack of very marked differences between diagnostic groups in their responses to any question in the SMCR indicates that a fairly large number of subjects will probably be required to establish reliable scoring weights.

Not only should the subjects in both criterion and cross validation populations be very competently diagnosed, but it would be valuable to have them evaluated psychiatrically on a wide variety of variables. In addition to nosological classification, the subjects could be rated upon significant personality traits, such as hostility or dependence, and upon dynamic factors such as sibling rivalry or parental rejection. Then, the same subject population could be regrouped accordingly when studying each different psychopathological variable. A prognostic index could also be developed by reconsidering the data after the progress of the subjects had been observed.

2. *Stimuli*

Further use of the Rorschach inkblots as stimuli in preference to other inkblots would seem inadvisable. As the present study has shown, there is little correlation with respect to the psychopathological significance of responses between a priori judgments based on the free association Rorschach and the results of the multiple choice Rorschach. Therefore, the large experiential background which has been built up with these particular inkblots under free association test conditions would be of little value in the construction of a multiple choice test. As a disadvantage, the multiple choice Rorschach provides suggestions to the subjects by listing possible responses which may influence their performance on a subsequent free association Rorschach.[2]

tology might be diagnosed differently at different installations.

A further attempt to validate the SMCR on schizophrenics by comparing the scores of subjects who had been discharged with those who still remained in the hospital was abandoned when it was found that in a population of the "better" schizophrenics, discharge was determined as much by the patient's home environment as by his psychiatric condition.

[2] For this reason, the scoring stencils for the SMCR have not been included in the present report. It was felt that the availability of these

While no quantitative evidence is available, it was the experience of the writer and of several other clinical psychologists who tested subjects individually subsequent to the SMCR that the performance of these subjects had been influenced by the preceding multiple choice Rorschach. The evaluation of the subjects' free association protocols was hindered by the difficulty in assessing the direction and extent of the influence of the multiple choice test.

Another advantage in using other than the Rorschach inkblots is the opportunity to construct inkblots particularly suited for a multiple choice test. Such blots could be structured to emphasize responses to either large areas or the blot as a whole. In the present experiment, the subjects rarely studied the slides minutely, but based their responses on a broad general impression. (Of course, it is possible that a more detailed examination of the stimuli would have been made had the subjects been more highly motivated or had the inkblots been presented in booklet form.) The inkblots could be structured to emphasize certain determinants, such as Levy's graduated "movement" series (61), or they might present a choice between determinants such as by containing two possible percepts, one form determined and the other predominantly color determined. It may be desirable to draw inkblot designs with printer's ink, as was done by Terman and Miles (52), rather than to use actual inblots. While it would be desirable to explore the potentialities of indefinite figures placed on simple backgrounds, it is very possible that better results are obtained when both figure and ground are indefinite, as with inkblots. This reservation should be kept in mind for all types of projective stimuli.

In addition to inkblots, various other projective stimuli should be investigated. Here again, it would not appear advisable to use the same stimuli as in the free association tests, but rather to develop new stimuli which together with the multiple choice responses could be structured in the direction of a specific psychopathological variable. For example, rather than employing the TAT cards, special pictures could be drawn. To investigate the maternal relationship, a picture of a mother and a child, the former with her arm upraised, might be shown. The responses could include one in which the mother is petting the child, another in which she is about to strike the child, and a third in which she is waving to someone behind the child. Of course, the picture would have to be sufficiently vague so that the subjects' responses would depend upon their projections. With the Rosenzweig Picture Frustration Test (43, 44) which is already structured to measure a specific variable, that of frustration tolerance, it might be necessary only to develop a set of multiple choice responses for each of the pictures.

Geometric figures, such as those employed in the Bender Gestalt (1) and the Franck Symbolic Sex Preferences (11) could be used. Here, the alternatives might consist of other geometric figures, with the subjects required to select the figure most closely resembling the stimulus. Perhaps particularly well suited for the detection of paranoid tendencies would be the Tautophone (13) in which a recording of low-voiced nonsense syllables is played. Subjects are asked what the recording is saying, being given to understand that it is producing a meaningful statement.

Further work can be done with story plots along the lines already developed by Kelly, and Bishop, Taylor and Wolfenstein (27, 50, 59). Verbal stimuli could also be confined to single words in a variation of the word association approach. Either fragmented letters or very brief exposures during which only one or two letters would stand out might be employed, with subjects being asked which word had been displayed.

3. *Administration of the Test*

It would seem desirable to present the stimuli as tests of ability, and in fact, to include some questions with obviously correct responses to further this belief. The best exposure time would have to be worked out separately for each different stimulus. While the results of the present experiment indicated that the length of exposure of the Rorschach slides made little difference in the amount of discrimination obtained, no conclusion can be reached as to whether this will hold true for other stimuli or for other psychopathological variables which may be investigated.

It may be possible to diminish the effect of suggestion and rationalization in the selection of responses and thereby obtain more significant projections by removing the stimulus before the multiple choices are presented to the subjects. For example, in the SMCR this could be done by listing the percepts in a question on a second slide which would be projected on the screen after the stimulus had been removed. The subjects could indicate their choice of response by marking the appropriate column on an IBM answer form. With this technique, the subjects

---

stimuli might lead to an indiscriminate use of the test, thus exposing many persons to suggested responses with the Rorschach blots. The scoring stencils will, of course, be made available to any qualified psychologist who wishes to perform a confirmatory or further exploratory experiment.

might be forced to reply more upon their initial free associative responses to the stimuli.

4. *Structure of the Questions*

No definitive conclusions could be reached from the results of the present study regarding the structure of the questions. Similar results were obtained when the subjects made a relative decision by selecting one of four percepts and when they made an absolute decision by rating a percept on an absolute five point rating scale. Differences due to variations in question "design" (that is, the types of percepts contained in a question) could not be investigated as originally planned because the psychopathological significance of the percepts had been determined on an a priori basis, and there proved to be little correlation between a priori judgment and the empirical results. However, little difference was found between "forced choice" questions in which subjects had to select a specific percept, and "unlimited choice" questions in which they could check the alternative "Something else."

Because of the lack of relationship between a priori judgment and empirical findings, it would seem advisable in subsequent experimentation to determine first, insofar as is possible, the psychopathological significance of each response empirically from a simple three or four point rating scale. Then, after the significance has been established empirically, the percepts could be fitted into the structures of different question designs. These designs could extend from the relatively simple one suggested by Malamud and Malamud (32) in which the subjects select either one of two percepts listed or the alternative "neither," to a design with as many as 10 percepts, as in the Harrower Multiple Choice Rorschach (19).

5. *Length of Test*

It would seem desirable to have the test as long as possible. The present experiment demonstrated that the amount of discrimination increased with the length of the test up to at least 160 questions, and it may very well continue beyond that point. With the use of a variety of stimuli rather than a repetition of the same 10 inkblots, the subjects should not become bored or fatigued as quickly. Moreover, it was felt that the discriminating power of a stimulus decreased as additional questions were applied to that stimulus. Therefore, using a higher proportion of stimuli to questions than in the SMCR may also serve to increase the validity.

6. *Scoring of the Test*

Since the five weighting systems employed in the present investigation all produced about the same amount of discrimination, it would not seem to make much difference which weighting system were to be used. However, in addition to the simple weighting of items, it may be advisable to make a Guttman scale analysis (16) of the responses. Few psychopathological variables are "pure" entities. This is particularly true of nosological categories. For example, psychiatrists now prefer to speak of the "schizophrenias" rather than of "schizophrenia," believing that this diagnosis encompasses a variety of only loosely associated conditions. Similarly, many divergent psychopathological factors may lead to the symptomatic diagnosis of anxiety reaction. A scale analysis may disclose a relatively small combination of responses which are pathognomonic of a particular type of psychopathology but which do not materially affect the subjects' total score when all questions are weighted equally.

VII. CONCLUSIONS

I. Further exploration with the multiple choice projective test approach in the detection of psychopathology would seem warranted since:

A. A fair validity has ben obtained with a multiple choice projective test. When scoring-weights based on a sample of 100 subjects in each diagnostic category were cross validated on a new sample of 50 subjects in each category, the percentage of subjects incorrectly identified by the SMCR was at most 30% for the normal-anxiety reaction comparison, 25% for the normal-schizophrenic comparison, and 30% for the anxiety reaction-schizophrenic comparison. There was an error of under 17% in predicting which of the three diagnostic categories a subject would not be in.

B. Split half reliability coefficients, when corrected by the Spearman-Brown prophecy formula, were .77 in the normal-anxiety reaction comparison, .87 in the normal-schizophrenic comparison, and .72 in the anxiety reaction-schizophrenic comparison.

C. There are grounds for believing that even greater validity may be obtained by:

(1). Further experimentation with various multiple choice projection tests.

(2). Combining a multiple choice projective test with a questionnaire and an intelligence test in a screening battery.

II. Of the different test conditions varied in the present experiment, only the length of the test was shown to be significantly related to the amount of discrimination obtained.

III. No significant differences in the amount of discrimination could be established from:

A. The application of different methods of weighting responses.

B. Increasing the size of the criterion population upon which the weights were developed from 100 to 150 subjects in each diagnostic category.

C. The subjects responding by making a "relative" as opposed to an "absolute" decision.

D. The subjects being forced to choose among specific percepts in a question as opposed to being given the "unlimited" alternative, "Something else."

E. Exposing the stimuli for a brief period as opposed to a lengthy period.

IV. There was little correlation between a priori judgments based on the free association Rorschach and the empirical results obtained with the SMCR with respect to the psychopathological significance of responses.

V. Both age and education, the latter perhaps a reflection of intelligence, may affect the subjects' responses.

VIII. APPENDIX

A. The Singer Multiple Choice Rorschach

Test A

NAME_____ AGE_____ GRADE_____ DATE_____

— 1 —

_____A wolf's head
_____An emblem
_____A topographical map
_____An octopus

— 2 —

_____Ruins of a castle
_____A blot
_____A squashed bug
_____An X-ray picture

— 3 —

_____An animal split open
_____Line down the center
_____A butterfly
_____Something else

— 4 —

_____A bomb bursting
_____Lungs and blood
_____Two dogs
_____Something else

— 5 —

_____Animal heads
_____Torn up butterfly
_____A map
_____Bears' heads

— 6 —

_____Two equal sides
_____Black and red paint
_____Dark clouds
_____A large eye

— 7 —

_____Two red roosters
_____Ribs
_____A fish
_____Lungs and chest

— 8 —

_____Vertebra
_____X-ray of the hips
_____Person tearing someone apart
_____Blood and black

— 9 —

_____Smoke
_____A hair ribbon
_____Nothing at all
_____Something else

— 10 —

_____Flag of a country
_____Mountain range
_____A black waterhole
_____Spilled ink

— 11 —

_____A cow with fur
_____Two feet
_____An X-ray
_____Something else

— 12 —

_____Necks of geese
_____Two scottie dogs
_____A spinal column
_____Sex organs

— 13 —

_____A pair of antlers
_____A butterfly with torn wings
_____Dried blood
_____Something else

— 14 —

_____Legs
_____A horrible object
_____A map
_____A butterfly

— 15 —

_____A line down the center
_____A bug
_____A black mess
_____Nothing at all

— 16 —

_____A spattered mess
_____A totem pole
_____A backbone
_____A turtle

— 17 —

_____An oil explosion
_____A bad design
_____A cut up male sex organ
_____Black paint

— 18 —

_____An insect
_____Jack the Ripper
_____A riverbed
_____Something else

— 19 —

_____Intestines
_____Ice formation
_____A squashed frog
_____A topographical map

— 20 —

_____Two heads
_____A coastline
_____Nothing at all
_____Something else

— 21 —

_____A butterfly
_____Grass
_____A clown's face
_____People watching each other

— 22 —

_____Two bears climbing
_____A biology slide
_____A sailboat
_____Something else

— 23 —

_____Head of small dog
_____Two polar bears climbing
_____An Easter egg
_____Modern art

— 24 —

_____Colored butterflies
_____An X-ray picture
_____An airplane
_____A blotch of colors

— 25 —
- Flowers
- Colored clouds
- Birth of a child
- Sea horses

— 26 —
- A dirty mess
- Smoke rising
- A bridge in the distance
- Nothing at all

— 27 —
- A person cut up
- A spinal column
- A halloween mask
- Something else

— 28 —
- Male sex organs
- Leaves
- Bugs
- People hanging in air

— 29 —
- Red and blue and yellow
- Bones
- Two spiders
- Something else

— 30 —
- A blue feathery flower
- A terrible germ
- Spiders attacking somebody
- Deer running

Test B

NAME_____ AGE_____ GRADE_____ DATE_____

— 1A —
- A bat
- Female sex organs
- A staring mask
- Something else

— 2A —
- Burnt piece of paper
- A spine
- A cat's face
- Two people

— 3A —
- A coastline
- A dark cloud
- A leaf
- Death

— 4A —
- Fire
- Two people
- A pair of pliers
- Smoke

— 5A —
- Human chest with a hole in it
- A squashed bug
- Blood
- A coastline

— 6A —
- Entrance to a tunnel
- Nothing at all
- A white lamp
- Something else

— 7A —
- Sex organs
- A skeleton effect
- Two staring eyes
- Picture from a medical book

— 8A —
- Waiters bowing
- A symmetrical design
- Animal cut open
- Something else

— 9A —
- Two devils
- Two monkeys hanging down
- A spinal cord
- Birds' heads

— 10A —
- A giant walking towards you
- Two snakes
- A furry man
- Something else

— 11A —
- A dragon skin
- A man sitting down
- A gorilla
- Snakes

— 12A —
- A relief map
- Clouds
- A torn chest
- Night

— 13A —
- Head of a mouse
- Profile of a man
- Antennae
- A dancer

— 14A —
- Two buffaloes crushing a rabbit
- Two equal sides
- Night
- Islands

— 15A —
- Dark smoke
- A bat
- The nervous system
- Something else

— 16A —
- A cat
- An X-ray
- A squashed worm
- Water

— 17A —
- Nothing at all
- Picture from an airplane
- Flying geese
- Something else

— 18A —
- Head of a dog
- A dragon fly
- A cloud
- A furry snake

— 19A —
- A Pacific island
- Two women go siping
- Male sex organs
- Something else

— 20A —
- A rabbit's head
- Clouds
- Scottie dogs
- Small dots on an island

— 21A —
- Smoke
- Just an inkblot
- A large eye
- Fog or mist

— 22A —

_____A spinal column
_____Decorated Christmas tree
_____Flowers
_____Ribs

— 23A —

_____Pink, blue, orange
_____Snow covered mountain
_____Male sex organs
_____A mountain

— 24A —

_____Two fish
_____Illustration in a medical book
_____A coat of arms
_____Something else

— 25A —

_____A famous painting
_____Staring eyes
_____Red, green, orange
_____Drawing in a medical book

— 26A —

_____The liver
_____Blood
_____Two women
_____Something else

— 27A —

_____Horns of a deer
_____An explosion
_____Intestines
_____A man's face

— 28A —

_____Squashed insects
_____A wishbone
_____Snakes eating a rabbit's head
_____Something else

— 29A —

_____A tobacco pipe
_____Two spiders
_____Deer running
_____Mountain sides

— 30A —

_____A Christmas tree
_____Modern art
_____Bloody insides
_____Illustration of different flowers

Test C

NAME _____ AGE _____ DATE _____

	EXACTLY LIKE IT	VERY MUCH LIKE IT	SOMEWHAT LIKE IT	ONLY A LITTLE LIKE IT	NOT AT ALL LIKE IT
— I —					
1. A pelvis					
2. Two Santa Clauses					
3. Nothing at all					
4. A bat					
5. Female sex organs					
6. Woman with her arms raised					
7. A staring mask					
8. Small islands					
9. A squashed bug					
10. Ruins of a castle					
— II —					
11. A man with his head cracked open					
12. A butterfly					
13. Two clowns playing pattycake					
14. Nothing at all					
15. Entrance to a tunnel					
16. Faces on the edges					
17. Two equal sides					
18. An arrowhead					
19. A large eye					
20. A white lamp in the center					
— III —					
21. A person tearing someone else apart					
22. A bloody mess					
23. A coastline					
24. A symmetrical design					
25. Waiters bowing					
26. Men picking something up					
27. A bow tie					

	EXACTLY LIKE IT	VERY MUCH LIKE IT	SOMEWHAT LIKE IT	ONLY A LITTLE LIKE IT	NOT AT ALL LIKE IT
28. A butterfly					
29. Vertebra					
30. Nothing at all					
— IV —					
31. Head of an animal					
32. A furry man					
33. Two snakes					
34. A skin stretched out tight					
35. A mountain range					
36. A black waterhole					
37. Nothing at all					
38. A fur rug					
39. A giant walking towards you					
40. Two boots					
— V —					
41. The nervous system					
42. A line down the center					
43. A rabbit					
44. A smashed body					
45. Dark smoke					
46. A bat					
47. An insect					
48. A mountain range					
49. Nothing at all					
50. A nutcracker					
— VI —					
51. Nothing at all					
52. Leg of a table					
53. A snake with feathers					
54. Picture from an airplane					
55. Black paint					
56. A cat's whiskers					
57. Flying geese					
58. A butterfly					
59. An oil explosion					
60. Rug of animal skin					
— VII —					
61. Two women gossiping					
62. A topographical map					
63. Grass					
64. Male sex organs					
65. A white lamp					
66. Two female faces					
67. A Pacific island					
68. Nothing at all					
69. A necklace					
70. Rocks balancing					
— VIII —					
71. Animals climbing					
72. A coat of arms					
73. Nothing at all					
74. Two fish					

	EXACTLY LIKE IT	VERY MUCH LIKE IT	SOMEWHAT LIKE IT	ONLY A LITTLE LIKE IT	NOT AT ALL LIKE IT
75. Blood					
76. A blotch of colors					
77. An evergreen tree					
78. Frogs' heads					
79. Colored butterflies					
80. Modern art					
— IX —					
81. Nothing at all					
82. The liver					
83. A burning candle					
84. Blood					
85. Two men in dunce caps					
86. Two women					
87. Smoke rising					
88. Bloody hands					
89. A colored map					
90. Lobsters					
— X —					
91. A biology slide					
92. Leaves					
93. A wishbone					
94. Nothing at all					
95. Snakes eating a rabbit's head					
96. Flowers					
97. Two lions					
98. Two people hanging in air					
99. A rabbit's head					
100. Squashed insects					

B. Six-Item Questionnaire to Screen Normals

Answer each question YES or NO. If you are not sure, guess.

1. Have you ever had treatment for a nervous condition?
2. In general, do you consider yourself a nervous person?
3. While in the service, were you ever hospitalized for psychoneurosis, combat fatigue, nervousness or some similar condition?
4. Since you have gotten out of the service, have you generally (not merely occasionally) felt irritable or jumpy?
5. Do you get upset more easily than the average person?
6. Have you ever filed a disability claim or collected a pension for psychoneurosis?

IX. BIBLIOGRAPHY

1. BENDER, LAURETTA. A visual motor Gestalt test and its clinical use. *Res. Monogr. Amer. orthopsychiat. Assn.*, 1938, No. 3.
2. BLAIR, G. M. & HOEHN, A. J. The use of the Multiple Choice Rorschach test with adolescents, *J. educ. Res.*, 1947, **41**, 297-304.
3. BRODY, M. B. Mental Testing. *J. ment. Sci.*, 1944, **90**, 127-151.
4. CHALKE, F. R. C. The Harrower Stress Tolerance Test. *Psychosom. Med.*, 1946, **8**, 215-216.
5. CLARK, R. M. A method of administering and evaluating the Thematic Apperception Test in group situations. *Genet. Psychol. Monogr.*, 1944, **30**, 3-55.
6. CROWN, S. A Controlled Association Test as a measure of neuroticism. *J. Personality*, 1947, **16**, 198-208.
7. DERI, SUSAN K. *Introduction to the Szondi Test*, New York: Grune and Stratton: 1949.
8. DORKEY, MARGARET, & AMEN, ELIZABETH W. A continuation study of anxiety reactions in young children by means of a projective technique. *Genet. Psychol. Monogr.*, 1947, **35**, 139-186.
9. DUE, F. O., WRIGHT, M. E., & WRIGHT, B. A. The Multiple Choice Rorschach test in military psychiatric differentiation: The use of statistical criteria. (In Reference 19, *Large Scale Rorschach Techniques and Multiple Choice Test*. Springfield, Ill., Charles C Thomas, 1945, Section V).
10. ELLIS, A., & CONRAD, H. S. The validity of personality inventories in military practice. *Psychol. Bull.*, 1948, **45**, 385-426.
11. FRANCK, K. Preferences for sex symbols and their personality correlates. *Genet. Psychol. Monogr.*, 1946, **33**, 73-123.
12. FRANK, L. K. The projective methods for the study of personality. *J. Psychol.*, 1939, **8**, 389-413.
13. GRINGS, W. W. The verbal summator technique and abnormal mental states. *J. abnorm. soc. Psychol.*, 1942, **37**, 529-545.
14. GUILFORD, J. P. *Fundamental statistics in psychology and education*. New York: McGraw-Hill Book Co., 1942.
15. GUILFORD, J. P. Some lessons from aviation psychology. *Amer. Psychol.*, 1948, **3**, 3-11.
16. GUTTMAN, L. A basis for scaling qualitative data. *Amer. Sociol. Rev.*, 1944, **9**, 139-150.
17. HARROWER, MOLLIE R., & GRINKER, R. R. The Stress Tolerance Test. Preliminary experiments with a new projective technique. *Psychosom. Med.*, 1946, **8**, 3-15.
18. HARROWER-ERICKSON, MOLLIE R. Large scale investigations with the Rorschach. *J. Consult. Psychol.*, 1943, **7**, 120-127.
19. HARROWER-ERICKSON, MOLLIE R., & STEINER, M. E. *Large scale Rorschach techniques and Multiple Choice Test*. Springfield, Ill., Charles C Thomas, 1945.
20. HARROWER-ERICKSON, MOLLIE R., & STEINER, M. E. Modification of the Rorschach method for use as a group test. *J. Genet. Psychol.*, 1943, **62**, 119-133.
21. HERTZ, MARGUERITE, ELLIS, A., & SYMONDS, P. M. Rorschach Methods and other projective techniques. *Rev. educ. Res.*, 1947, **17**, 78-100.
22. HERTZMAN, M. A comparison of the individual and group Rorschach tests. *Ror. Res. Exch.*, 1942, **6**, 89-108.
23. HOROWICZ, RUTH E. A pictorial method for the study of self-identification in preschool children. *J. Genet. Psychol.*, 1943, **62**, 135-148.
24. HUNT, W. D., & STEVENSON, I. Psychological testing in military clinical psychology. II Personality testing. *Psychol. Rev.*, 1946, **53**, 107-115.
25. HUSTON, P. E., & SHAKOW, D. Studies of motor function in schizophrenia. III Steadiness. *J. Gen. Psychol.*, 1946, **34**, 119-126.
26. JENSEN, M. B., & ROTTER, J. B. The value of thirteen psychological tests in officer candidate screening. *J. Appl. Psychol.*, 1947, **31**, 312-322.
27. KELLY, G. A. & BISHOP, F. A projective method of personality investigation. *Psychol. Bull.*, 1942, **39**, 599.
28. LAWSHE, C. H., JR., & FORSTER, M. H. Studies in projective techniques: I The reliability of a multiple choice group Rorschach test. *J. Appl. Psychol.*, 1947, **31**, 199-211.
29. LINDNER, R. M. A further contribution to the group Rorschach. *Ror. Res. Exch.*, 1943, **7**, 2-15.
30. LINDNER, R. M., CHAPMAN, K. W., & RINCK, E. C. The development of a group Rorschach technique in a federal penal institution with special reference to the problem of psychopathic personality. *Psychol., Bull.*, 1942, **39**, 513-514.
31. MALAMUD, D. I. Value of the Maller Controlled Association Test as a screening device. *J. Psychol.*, 1946, **21**, 37-43.
32. MALAMUD, RACHEL F., & MALAMUD, D. I. Validity of the amplified multiple choice Rorschach as a screening device. *J. Consult. Psychol.*, 1945, **9**, 224-227.
33. MALAMUD, RACHEL F., & MALAMUD, D. I. The multiple choice Rorschach: a critical examination of its scoring system. *J. Psychol.*, 1946, **21**, 237-242.
34. MALLER, J. B. *Controlled Association Test*

(Forms A and B). New York: Bureau of Publications, Teachers College, Columbia Univ. 1934.

35. MICHAEL, J. C., & BUHLER, C. Experiences with personality testing in a neuropsychiatric department of a public general hospital. *Dis. nerv. Syst.*, 1945, **6**, 205-211.
36. MITTLEMANN, B., & BRODMAN, K. The Cornell Indices and the Cornell Word Form I. Construction and Standardization *Ann. N.Y. Acad. Sci.*, 1946, **46**, 573-577.
37. MUNROE, RUTH. Prediction of the adjustment and academic performance of college students by a modification of the Rorschach method. *Appl. Psychol. Monogr.*, Stanford Univ. Press, 1945, **7**, pp. 104.
38. MURRAY, H. A., & STEIN, M. A note on the selection of combat officers. *Psychosom. Med.*, 1943, **5**, 386-391.
39. PRESSEY, S. L. A group scale for investigating the emotions. *J. abnorm. soc. Psychol.*, 1921, **16**, 55-64.
40. PRESSEY, S. L., & PRESSEY, L. W. "Cross-Out" tests with suggestions as to a group scale of the emotions. *J. Appl. Psychol.*, 1919, **3**, 138-150.
41. RAPAPORT, D. The Szondi Test. *Bull. Menninger Clin.*, 1941, **5**, 33-39.
42. RAPAPORT, D., & SCHAFER, R. The Rorschach test: a clinical evaluation. *Bull. Menninger Clin.*, 1945, **9**, 73-77.
43. ROSENZWEIG, S. The picture association method and its application in a study of reactions to frustration. *J. Personality*, 1945, **14**, 3-23.
44. ROSENZWEIG, S., CLARKE, H. J., GARFIELD, M. S., & LEHNDORFF, A. Scoring samples for the Rosenzweig Picture-Frustration Test. *J. Psychol.*, 1946, **21**, 45-72.
45. ROSS, W. D., & ROSS, S. Some Rorschach ratings of clinical value. *Ror. Res. Exch.*, 1944, **8**, 1-9.
46. SARGENT, HELEN. Projective Methods: their origins, theory and application in personality research. *Psychol. Bull.*, 1945, **42**, 257-293.
47. SPRINGER, N. N. The validity of the multiple choice Rorschach in the screening of naval personnel. *J. Gen. Peychol.*, 1946, **35**, 27-32.
48. SYMONDS, P. M., & KRUGMAN, M. Projective methods in the study of personality. *Rev. educ. Res.*, 1944, **14**, 81-98.
49. SYMONDS, P. M., & SAMUEL, E. A. Projective methods in the study of personality. *Rev. educ. Res.*, 1941, **11**, 80-93.
50. TAYLOR, W. S. Reality and fantasy thinking in social reconstruction. *Indian J. Soc. work.*, 1946, **7**, 169-175 (abstract).
51. TEMPLE, RITA, & AMEN, ELIZABETH W. A study of anxiety in young children by means of a projective technique. *Genet. Psychol. Monogr.*, 1944, **30**, 59-113.
52. TERMAN, L. M., & MILES, CATHERINE C. *Sex and personality (Studies in masculinity and femininity).* New York: McGraw-Hill Book Co., 1936.
53. WEIDER, A., BRODMAN, K., MITTLEMANN, B., WECHSLER, D., & WOLFF, H. G. The Cornell Index: a method for quickly assaying personality and psychosomatic disturbances to be used as an adjunct to interview. *Psychosom. Med.*, 1946, **8**, 411 413.
54. WEIDER, A., MITTLEMANN, B., WECHSLER, D., & WOLFF, H. G. The Cornell Selective Index. *J. Amer. Med. Assn.*, 1944, **124**, 224-228.
55. WEIDER, A., & WECHSLER, D. The Cornell Indices and the Cornell Word Form. 2. Results. *Ann. N.Y. Acad. Sci.*, 1946, **46**, 579-591.
56. WEIDER, A. *et al. Cornell Index Manual.* New York: The Psychological Corporation, 1946.
57. WHITE, R. W. Interpretation of Imaginative Productions. (From Hunt, J. McV. *Personality and the behavior disorders,* New York: Ronald Press, 1944. Vol. I, Chapt. VI.)
58. WITTMAN, P. The multiple choice Rorschach test as a differential diagnostic tool. *Arch. Neurol. Psychiat.* Chicago, 1946, **55**, 154-156.
59. WOLFENSTEIN, M. The reality principles in story preferences of neurotics and psychotics. *Character & pers.*, 1944, **13**, 135-151.
60. ZUBIN, J. Recent advances in screening the emotionally maladjusted. *J. Clin. Psychol.*, 1948, **4**, 56-63.
61. ZUBIN, J., LEVY, D. M., & RUST, R. Movement responses in normals, schizophrenics and neurotics on the Levy Movement Cards. *Amer. Psychol.*, 1947, **2**, 269.